

PRINCIPAL COMPONENT ANALYSIS- BASED COMBUSTION MODELS

by

Benjamin J. Isaac

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemical Engineering

The University of Utah

May 2014

Copyright © Benjamin J. Isaac 2014

All Rights Reserved

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of Benjamin J. Isaac
has been approved by the following supervisory committee members:

<u>Philip J. Smith</u>	, Chair	<u>3/12/2014</u> Date Approved
<u>Alessandro Parente</u>	, Member	<u> </u> Date Approved
<u>Gérard Degrez</u>	, Member	<u> </u> Date Approved
<u>James Sutherland</u>	, Member	<u>3/12/2014</u> Date Approved
<u>Jeremy Thornock</u>	, Member	<u>3/12/2014</u> Date Approved

and by Milind Deo, Chair/Dean of
the Department/College/School of Chemical Engineering

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Energy generation through combustion of hydrocarbons continues to dominate as the most common method for energy generation. In the U.S., nearly 84% of the energy consumption comes from the combustion of fossil fuels. Because of this demand, there is a continued need for improvement, enhancement, and understanding of the combustion process. As computational power increases, and our methods for modelling these complex combustion systems improve, combustion modelling has become an important tool in gaining deeper insight and understanding of these complex systems. The constant state of change in computational ability leads to a continual need for new combustion models that can take full advantage of the latest computational resources. To this end, the research presented here encompasses the development of new models which can be tailored to the available resources, allowing one to increase or decrease the amount of modelling error based on the available computational resources and desired accuracy. Principal component analysis (PCA) is used to identify the low-dimensional manifolds which exist in turbulent combustion systems. These manifolds are unique in their ability to represent a larger dimensional space with fewer components, resulting in a minimal addition of error. PCA is well-suited for the problem at hand because of its ability to allow the user to define the amount of error in approximation, depending on the resources at hand. The research presented here looks into various methods which exploit the benefits of PCA in modelling combustion systems, demonstrating several models, and providing new and interesting perspectives for the PCA-based approaches to modelling turbulent combustion.

CONTENTS

ABSTRACT	iii
LIST OF FIGURES	vi
LIST OF TABLES	xi
ACKNOWLEDGMENTS	xii
CHAPTERS	
1. INTRODUCTION	1
2. A NOVEL METHODOLOGY FOR CHEMICAL TIME-SCALE EVALUATION WITH DETAILED CHEMICAL REACTION KINETICS	10
2.1 Abstract	10
2.2 Introduction	11
2.3 Methodology	16
2.3.1 Principal component analysis	16
2.3.2 Principal variables	17
2.3.3 Jacobian matrix down-sizing	18
2.4 Test cases	20
2.5 Results and discussion	22
2.6 Conclusions	42
2.7 Acknowledgments	42
3. REDUCED-ORDER PCA MODELS FOR CHEMICAL REACTING FLOWS	45
3.1 Abstract	45
3.2 Introduction	46
3.3 Principal component analysis	47
3.4 Challenges of the MG-PCA model	51
3.4.1 Reference data-set	51
3.4.2 Computation of the B matrix	52
3.4.3 Selection of the transported variables	54
3.4.4 Centering and scaling	58
3.4.5 Comparison of the PC-score approach, MG-PCA, and MG-L-PCA	60
3.4.6 MG-L-PCA	63
3.5 Results	66

3.5.1	Auto-ignition delay time	66
3.5.2	Flame-turbulence interaction	67
3.6	Conclusions	70
3.7	Acknowledgments	74
4.	MODELLING COMBUSTION SYSTEMS WITH PRINCIPAL COMPONENT ANALYSIS TRANSPORT EQUATIONS	76
4.1	Abstract	76
4.2	Introduction	77
4.3	Data-sets	78
4.4	Theory	80
4.4.1	Principal component analysis	80
4.4.2	Regression models	82
4.5	Results and discussion	84
4.5.1	A priori model evaluation	84
4.5.2	A posteriori model evaluation	93
4.6	Conclusion	109
4.7	Acknowledgments	109
5.	ADVANCING THE PCA APPROACH FOR COMBUSTION SYSTEMS	110
5.1	Introduction	110
5.2	PCCE	111
5.2.1	PCCE theory	111
5.2.2	PCCE results and conclusions	114
5.3	MF-PCA	116
5.3.1	MF-PCA theory	118
5.3.2	MF-PCA results and conclusions	121
5.4	Conclusions	123
6.	CONCLUSION	126
	REFERENCES	130

LIST OF FIGURES

1.1	A scatter plot of example variables X_1 (x-axis) versus X_2 (y-axis), illustrating the correlation between the variables.	3
1.2	Centered and scaled variables X_{1SC} (x-axis) versus X_{2SC} (y-axis) which have been centered by their means and scaled by their standard deviations.	3
1.3	Principal components Z_1 (x-axis) versus Z_2 (y-axis).	4
1.4	Principal components Z_1 (x-axis) versus Z_2 (y-axis) with the truncated versions of PCs indicated in green and red markers.	6
1.5	Centered and scaled variables X_{1SC} (x-axis) versus X_{2SC} (y-axis) with approximated versions using the truncated basis. The red markers show the truncation where the 2 nd component was removed, and the green markers show the truncation where the 1 st component was removed.	6
2.1	Flow diagram describing the process of analysis used to obtain the Damköhler number. Here the subscripts i, r, and pv denote species, reaction number, and principal variable, respectively.	19
2.2	Temperature as a function of mixture fraction ξ for cases HM1, HM2, and HM3.	21
2.3	Chemical time-scale τ_c (a), and Damköhler values Da_η (b) as a function of mixture fraction ξ . Full Jacobian matrix and unfiltered time-scales.	23
2.4	Chemical time-scale τ_c (a), and Damköhler values Da_η (b) as a function of mixture fraction ξ for the full Jacobian matrix. Time-scales above 1000 seconds have been removed.	24
2.5	Chemical time-scale τ_c (a), and Damköhler values Da_η (b) as a function of mixture fraction, ξ , using only the major species in the reduced Jacobian matrix.	26
2.6	Trace plot for cases HM1, HM2, and HM3. Y axis gives a normalized variance which is lost based on the selection of q, the x-axis.	27
2.7	Chemical time-scales associated to the full (black) and reduced (red) Jacobian matrix (red line). Reduced Jacobian obtained using 10 PVs, determined using method B2. HM1 case.	28
2.8	Chemical time-scale τ_c (a), and Damköhler values Da_η (b) as a function of mixture fraction, ξ . Jacobian matrix restricted to ten principal variables. ...	29
2.9	Da_η values as a function of the mixture fraction along the burner axis for flame HM1, varying the number of principal variables.	31

2.10	Damköhler numbers calculated using the integral and the Kolmogorov time-scales. The area bounded between the curves represents the range of Damköhler numbers between the largest mixing scales (Da_I) and the smallest mixing scales (Da_η). The dashed line represents a Damköhler number of one along the entire mixture fraction space. HM1 data-set.	32
2.11	Damköhler numbers calculated using the integral and the Kolmogorov time-scales. The area bounded between the curves represents the range of Damköhler numbers between the largest mixing scales (Da_I) and the smallest mixing scales (Da_η). The dashed line represents a Damköhler number of one along the entire mixture fraction space. HM2 data-set.	33
2.12	Damköhler numbers calculated using the integral and the Kolmogorov time-scales. The area bounded between the curves represents the range of Damköhler numbers between the largest mixing scales (Da_I) and the smallest mixing scales (Da_η). The dashed line represents a Damköhler number of one along the entire mixture fraction space. HM3 data-set.	34
2.13	Contours showing Da_η values along the radial and axial direction of the burner. Colors in the graphs represent Da_η and isolines show the temperature. Plots (a), (b), and (c) show the contours for the cases HM1, HM2, and HM3, respectively.	35
2.14	Stacked bar chart showing normalized chemical species weights from the eigenvector matrix corresponding to chemical time-scales from the inverse of the eigenvalues.	36
2.15	Comparison of chemical time-scales τ_c for the HM1 data-set calculated using the principal variable approach (black 'x' markers) or with global chemistry (blue '+' markers) and τ_c from the calculation of HM1 using global chemistry with the EDC model (red 'diamond' markers).	38
2.16	Chemical time-scale τ_C (a), and Damköhler Da_η (b) values as a function of mixture fraction (ξ) using the full Jacobian matrix.	39
2.17	Trace plot for DNS case. Y-axis gives a normalized variance which is lost based on the selection of q , the x-axis.	40
2.18	Chemical time-scale τ_C (a), and Damköhler Da_η (b) values as a function of mixture fraction (ξ), using two principal variables.	41
2.19	Damköhler numbers calculated using the integral time-scale and the Kolmogorov time-scale using two principal variables	43
3.1	Initial turbulence field for the 2D DNS, x-component velocity	53
3.2	Y_{H_2O} , Y_{HCO} field for the 2D DNS, $t = 7.11799 \cdot 10^{-4}$ s.	53
3.3	Maximum <i>nrms</i> distance (y-axis) for the reconstruction of the state-space variables as a function of the number of variables (q) (x-axis), while using the $\tilde{\mathbf{Z}}_q$ or \mathbf{Z}_q in the construction of the \mathbf{B} matrix. The dashed line shows 5% <i>nrms</i> . Scaling: standard deviation. PV selection: B2.	55
3.4	Lost variance (y -axis) while adding principal variables (x -axis). The trace is given for the various PV selection methods highlighted in Section 3.4. Scaling: standard deviation	57

3.5	Lost variance (y -axis) while adding principal variables (x -axis). The trace is given for the various scaling methods highlighted in Section 3.4. PV selection: B2.	61
3.6	Minimum R^2 statistic for the state variables as a function of the number of retained variables (x -axis) and clusters (y -axis). Scaling: pareto. Fuel: Syngas.	64
3.7	Minimum R^2 statistic for the principal variable source terms as a function of the number of retained variables (x -axis) and clusters (y -axis). Scaling: pareto. Fuel: Syngas	65
3.8	Auto-ignition delay time using MG-PCA as a function of temperature for various system pressures, while $q = 6$ (a) or $q = 7$ (b).	68
3.9	Auto-ignition delay time using MG-L-PCA as a function of temperature for various system pressures, while using $q = 4$ (a) or $q = 5$ (b).	69
3.10	Fields of Y_{H_2O} (top) and Y_{HCO} (bottom) using DNS (left) and MG-PCA (right) $t = 7.11799 \cdot 10^{-4}$ s.	71
3.11	Scatter plot of (from top to bottom) Y_{CO} , Y_H vs temperature, using DNS (left) and MG-PCA (right) $t = 7.11799 \cdot 10^{-4}$ s.	72
3.12	Scatter plot of (from top to bottom) $Y_{H_2O_2}$, Y_{HO_2} vs temperature, using DNS (left) and MG-PCA (right) $t = 7.11799 \cdot 10^{-4}$ s.	73
4.1	$Nrms$ error values for several major species (left), and minor species (right), while varying q , the number of PCs, and the scaling method.	85
4.2	$Nrms$ error values for s_{Z_1} while varying q , the number of PCs, and the scaling method.	87
4.3	Scree plot from the eigenvalue matrix, showing the fraction of explained variance (y -axis) as a function of the number of PCs (q) for the system containing a subset of the original species (' \mathbf{x} ' markers), and the fulls system (' \mathbf{o} ' markers).	91
4.4	A scatter plot of mixture fraction (x -axis) versus Z_1 (y -axis), illustrating the correlation between the variables.	92
4.5	PSR temperature as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q	94
4.6	Major species products as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q	95
4.7	Major species reactants as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q	96
4.8	Minor species as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q	97

4.9	Minor species as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q .	98
4.10	Minor species as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q .	99
4.11	A scatter plot of mixture fraction (x-axis) versus Z_1 (y-axis), illustrating the correlation between the variables.	103
4.12	Scatter plots representing the $q = 2$ dimensional manifold with Z_1 on the x-axis, Z_2 on the y-axis, and the gray-scale representing s_{Z_1} . Plot (a) shows the representation of the manifold from the training data-set, and the plot (b) shows the represented manifold from the simulation results.	104
4.13	Density (left) and temperature (right) fields from the ARCHES calculation at $t = 0.25$ seconds.	105
4.14	Z_1 (left) and Z_2 (right) fields from the ARCHES calculation at $t = 0.25$ seconds. Z_1 is highly correlated with Bilger's mixture fraction, and Z_2 is related to the extent of reaction being highly correlated with the temperature of the system.	105
4.15	s_{Z_1} (left) and s_{Z_2} (right) fields from the ARCHES calculation at $t = 0.25$ seconds.	106
4.16	CO (left) and H_2 (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.	106
4.17	CO_2 (left) and H_2O (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.	107
4.18	O_2 (left) and OH (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.	107
4.19	O (left) and H (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.	108
4.20	HCO (left) and HO_2 (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.	108
5.1	Total elemental mass fractions (K_e^T) for the flamelet data-set as a function of mixture fraction	113
5.2	K_O (left) and K_H (right) elemental mass fractions as a function of mixture fraction while varying the number of constraints.	115
5.3	K_C (left) and K_N (right) elemental mass fractions as a function of mixture fraction while varying the number of constraints.	115
5.4	Temperature (left) and density (right) as a function of mixture fraction while varying the number of constraints.	117
5.5	CO (left) and H_2 (right) mass fractions as a function of mixture fraction while varying the number of constraints.	117

5.6 OH radical (a) and conditioned OH radical (b) as a function of mixture fraction, with red markers representing the condition function $f_{OH}(\xi)$	119
---	-----

LIST OF TABLES

1.1	Smallest relevant time-scale (τ_a) for a DNS case while comparing the MG-PCA and PC-transport models.	8
3.1	<i>nrms</i> distance for the reconstructed state-space with $q = 7$ while testing various PV selection methods. The acronym <i>pv</i> indicates if the variable is a principal variable. Scaling method: auto scaling.	59
3.2	<i>nrms</i> distance for the reconstructed state-space using $q = 7$ variables, while testing various scaling methods. S = <i>auto</i> , R = <i>range</i> , P= <i>pareto</i> , V= <i>vast</i> , L= <i>level</i>	61
3.3	<i>nrms</i> distance and R^2 statistics for the reconstructed state-space with $q = 7$. Scaling: standard deviation. PV selection: B2.	62
3.4	<i>nrms</i> distance and R^2 statistics for the reconstructed source terms with $q = 7$. Scaling: standard deviation. PV selection: B2.	62
4.1	Manifold nonlinearity (χ_{Φ}) for state-space variables, Φ while using different scaling methods.	89
4.2	<i>Nrms error</i> and R^2 statistics for the prediction of s_{Z_1} while using pareto scaling and $q = 2$ or $q = 3$	89
4.3	<i>nrms error</i> and R^2 statistics for the prediction of Φ while using pareto scaling and $q = 2$	91
4.4	Eigenvector matrix, A, from the PC analysis.	92
4.5	<i>Nrms error</i> and R^2 statistics for the prediction of Φ for the laminar data-set while using pareto scaling and $q = 2$	102
4.6	Eigenvector matrix, A, from the PC analysis.	102
5.1	Fraction of total elemental mass remaining $1 - \sum K_e^{res}$ while varying the number of constraints on the system.	115
5.2	<i>nrms</i> and R^2 error for the reconstruction of the species mass fractions and s_{Z_1} using the standard PCA approach or MF-PCA	122
5.3	Eigenvector matrix, A, from the PC analysis.	124
5.4	<i>nrms</i> and R^2 error for the reconstruction of a reduced-set of species mass fractions and s_{Z_1} using the standard PCA approach or MF-PCA. Here a comparison is made when $q = 1$ for both approaches.	124
5.5	<i>nrms error</i> and R^2 statistics for the prediction of Φ while using pareto scaling and $q = 1$	124

ACKNOWLEDGMENTS

First and foremost, I wish to acknowledge and praise my wife Andrea for the unwavering support, love, concern, and strength she has given me during these last several years. This education and experience was made possible because of her willingness to sacrifice and put her own career ambitions on hold while I finished this degree. I am confident in saying that this great accomplishment is just as much hers as it is mine.

I wish to acknowledge my parents for their support and encouragement that they have continually given me. From a very young age, they taught me the value of hard work, and showed me how to set goals and achieve them.

Working with Phil has been an enlightening experience. The most important thing I have learned from Phil is the power of having confidence in others. The trust he had in me helped me to have the confidence that I needed to learn, grow, and work diligently, even while facing extremely difficult problems encountered during my research. Phil is incomprehensibly optimistic, and this is one of the reasons he is very successful. I am grateful for the time that I have had and will continue to have under his leadership.

Alessandro has had a monumental influence on me personally and through my education. I am very grateful for the patience he has had with me while learning and for the extreme care he has taken in guiding me and helping me to achieve my goals. Alessandro has the ability to inspire and to open the minds of others to new ideas and new concepts.

Finally, I would like to thank the supervisory committee for the time they spent in speaking with me, teaching me, and working with me during these years.

The research was sponsored by the National Nuclear Security Administration under the Acceleration Development of Retrofittable CO₂ Capture Technologies through Predictivity program through DOES Cooperative Agreement DE-NA0000740.

CHAPTER 1

INTRODUCTION

Energy generation through combustion of hydrocarbons continues to dominate as the most common method for energy generation. In the U.S., nearly 84% of the energy consumption comes from the combustion of fossil fuels [2]. Because of this demand, there is a continued need for improvement, enhancement, and even understanding of the combustion process. As computational power increases and our methods for modelling these complex combustion systems improve, combustion modelling has become an important tool in gaining deeper insights and understanding of these complex systems.

The constant state of change in computational ability leads to a continual need for new combustion models that can take full advantage of the latest computational resources. To this end, the research presented here encompasses an investigation into new models which can be tailored to the available resources, allowing one to increase or decrease the amount of modelling error based on the desired complexity of the model.

The body of the dissertation contains three publications which outline various aspects of the development and use of these models, as well as an additional chapter discussing some new concepts which are the basis for future work. In order to tie the concepts presented in the dissertation together, this introductory chapter will briefly summarize the work in these publications and provide the transitional discussion which is needed.

In order to develop a model which is adaptive to computation resources, a method for identifying the key characteristics of the system and associating an importance to these characteristics is needed. Recently, principal component analysis (PCA) began to receive some attention for its application within the turbulent combustion field. PCA is well-suited for this task as it analyzes the covariance between variables in a system and identifies a linear representation of the system through orthogonal vectors, with each vector having an associated level of importance.

The basic concepts of PCA are easily described in a 2D example. Figure 1.1 shows a scatter plot of the variables X_1 and X_2 . One can easily identify that the data have some

degree of correlation. In order to perform a PC analysis, first the data must be centered and scaled in order to compare the data evenly when computing the covariance matrix. In general, the variables are centered and scaled according to the equation:

$$X_{iSC} = (X_i - \bar{X}_i)/d_i \quad (1.1)$$

where X_{iSC} is the centered and scaled version of variable X_i , \bar{X}_i and d_i are the mean and scaling factor of X_i . In general, d_i is taken as the standard deviation of the variable. Figure 1.2 shows the centered and scaled versions of X_1 and X_2 . Given that the variables have been scaled to the same relative size, the covariance matrix (\mathbf{S}) of the data is now computed:

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}. \quad (1.2)$$

The covariance describes the correlations between variables, and is essential for identifying the underlying structure in a data-set. Next, the eigenvalue decomposition of \mathbf{S} follows yielding the eigenvectors and eigenvalues of the system:

$$\mathbf{S} = \mathbf{A} \mathbf{L} \mathbf{A}^T. \quad (1.3)$$

The eigenvalues (\mathbf{L}) in the system are interesting as they quantify the amount of variation described by each of the corresponding eigenvectors (\mathbf{A}). Now, given the basis matrix (\mathbf{A}), \mathbf{X} is now projected onto the new basis in order to compute the PCs of the system:

$$\mathbf{Z} = \mathbf{X} \mathbf{A}. \quad (1.4)$$

A visualization of the PCs, \mathbf{Z} , is given in Figure 1.3. It is important to note that until now, no truncation has been made on the system, yielding 2 PCs, one for each of the example variables. The figure shows that the variables appear to be orthogonal to each other, describing different degrees of variation. Next, a truncation is made on \mathbf{A} , leaving \mathbf{A}_q . This truncation reduces the number of PCs, leaving:

$$\mathbf{Z}_q = \mathbf{X} \mathbf{A}_q. \quad (1.5)$$

\mathbf{Z}_q is shown in Figure 1.4, showing the results from removing the first eigenvector (green markers) or the second eigenvector (red markers). Because the first eigenvector pertains to the majority of variance in the system (being that it corresponds to the first eigenvalue), the truncation of this eigenvector leads to the poorest truncation in PC space. When the second eigenvector is truncated, it is clear that the largest amount of variation in the system

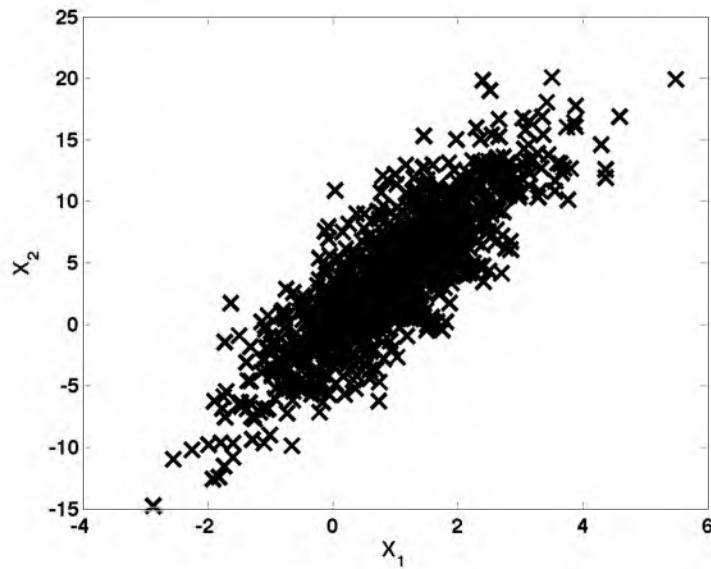


Figure 1.1: A scatter plot of example variables X_1 (x-axis) versus X_2 (y-axis), illustrating the correlation between the variables.

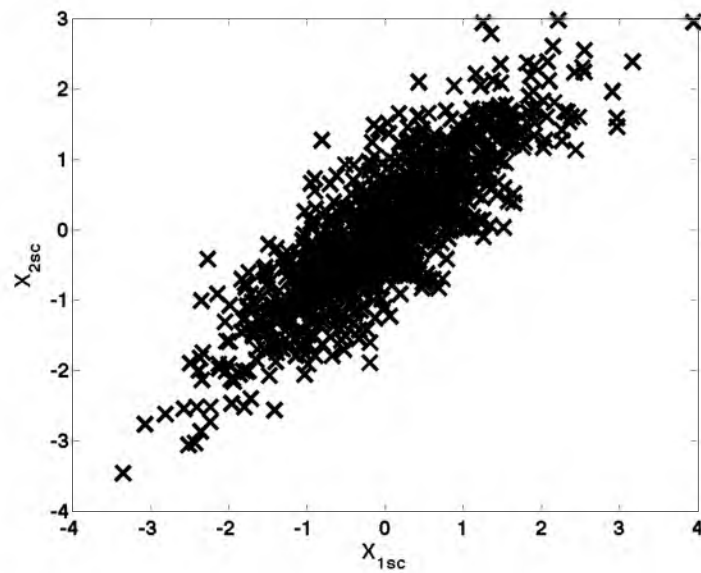


Figure 1.2: Centered and scaled variables X_{1sc} (x-axis) versus X_{2sc} (y-axis) which have been centered by their means and scaled by their standard deviations.

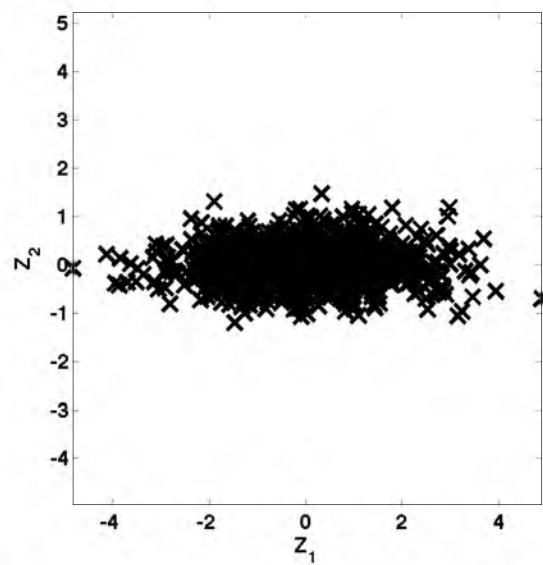


Figure 1.3: Principal components Z_1 (x-axis) versus Z_2 (y-axis).

remains. Now, \mathbf{Z}_q can be used to approximate the data. The following equation is used to approximate \mathbf{X} :

$$\mathbf{X} \approx \mathbf{Z}_q \mathbf{A}_q^T. \quad (1.6)$$

The resultant approximation is shown in Figure 1.5, using either of the truncation's described above. Truncation of the second eigenvector results in the best approximation, which is shown with the red markers. It is clear from the 2D example that PCA has the ability to reduce the dimensionality of the data by identifying the underlying structure of the system, and reducing in dimension by removing coordinates which describe smaller degrees of variance. These properties are attractive, when looking at systems such as combustion, where many of the variables can easily be described by the underlying structure or manifold that the system naturally exhibits, thus allowing for a reduction in modelling parameters.

The initial application of PCA to combustion systems has produced promising results. Parente [61] used PCA to identify the underlying behavior which governs the reaction and evolution of the chemical species for a syngas [80] and methane flame [79]. Sutherland and Parente take the PCA-based approach one step further by deriving transport equations for the principal components, creating the concept of a model with the unique properties of PCA [85]. The PC-transport model can then represent a system perfectly by transporting all of the PCs for the system, or one can reduce the number of PCs, introducing some degree of error (depending on the size of the reduction), and still maintain a good representation of the system. This concept fulfills the basic requirements of a model, which can tune the error in approximation based on the available computational resources and desired error. In addition, the PC analysis identifies which PCs are most important to transport, allowing one to make informed decisions for reducing the model while maintaining the largest degree of accuracy.

Because of PCA's unique ability in identifying the underlying structure or the low-dimensional manifold in a system, various additional methods have been created which utilize the basis derived from PCA, such as a principal variables analysis (PVA). PVA uses the basis matrix from PCA to identify the principal variables in a system, or the variables which contain most of the variation in a system. The PV analysis is very interesting as it helps to identify which variables are pertinent to the description of a system. An intuitive application of PCA and PVA to turbulent combustion can be seen in a time-scale analysis. In general, the time-scales of a system (τ_a) can be identified through an eigenvalue decomposition of the source-term Jacobian matrix [23]:

$$\tau_a = \frac{1}{|\lambda_a|} \quad (1.7)$$

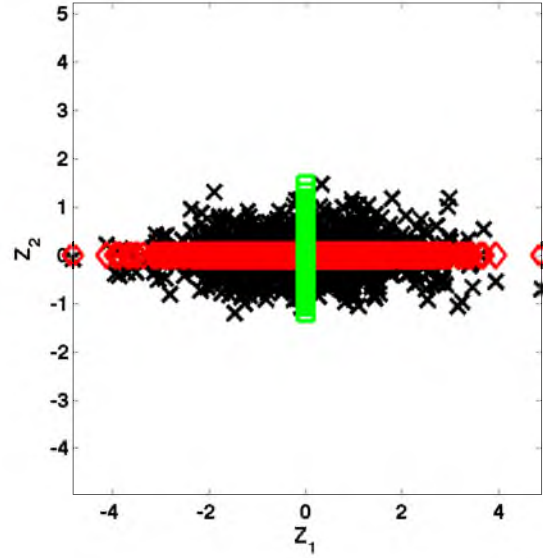


Figure 1.4: Principal components Z_1 (x-axis) versus Z_2 (y-axis) with the truncated versions of PCs indicated in green and red markers.

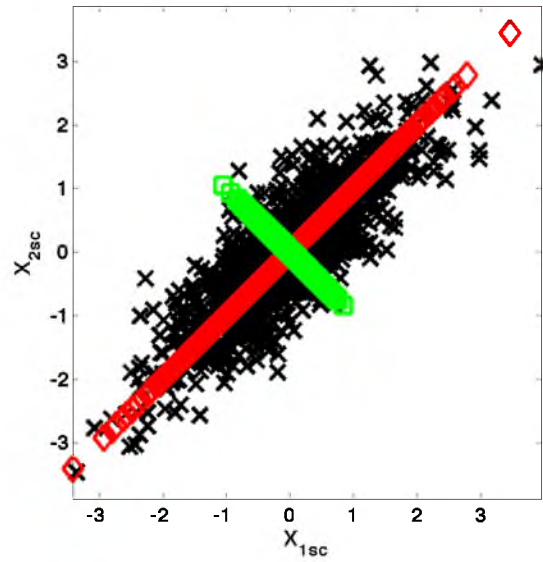


Figure 1.5: Centered and scaled variables X_{1sc} (x-axis) versus X_{2sc} (y-axis) with approximated versions using the truncated basis. The red markers show the truncation where the 2^{nd} component was removed, and the green markers show the truncation where the 1^{st} component was removed.

where λ_a is the set of eigenvalues from the decomposition. There is, however, an inherent noise associated with this calculation due to the dormant chemical species in the system. PVA can be used to identify the variables which are not dormant. When calculating the time-scales with the PVs of the system, the results can be much more clear. This concept is discussed thoroughly in the first publication: “A Novel Methodology for Chemical Time-Scale Evaluation with Detailed Chemical Reaction Kinetics” found in Chapter 2.

The PVs of the system contain most of the pertinent evolution in the system. Accordingly, Coussement et al. [16] developed a combustion model (MG-PCA) which transports the PVs of the system and use the PC basis, \mathbf{A} , in order to approximate the nontransported species. Chapter 3, “Reduced-order PCA Models for Chemical Reacting Flows” is the second publication, which describes the MG-PCA methods. A detailed description of the approach is given and some changes to the model are suggested in order to increase the accuracy for the model. With the changes in place, the model is tested in both *a priori* studies and demonstrated within numerical solvers.

The implementation of the MG-PCA approach is straightforward, making it attractive from an implementation perspective. However, use of the optimal basis calculated directly from PCA remains the most attractive option, especially when the method is combined with nonlinear regression [5, 72]. An important feature of PCA, when applied to combustion systems, can be seen in a time-scale analysis. Based on the type of scaling used in PCA, the leading PCs are, in general, heavily weighted with the major species in the system, which evolve more slowly. Table 1.1 shows a comparison of the smallest relevant time-scales calculated from simulation results comparing DNS, the MG-PCA approach, and the standard PC-Transport approach (without regression). The simulations are for a premixed syngas, and H_2 jet running at stoichiometric conditions with unity Lewis number [18]. Here it is observed that the PCA approach, in general, reduces the stiffness of the system. Because of this, there appears to be a greater advantage in directly transporting the PCs of the system, especially if the system can be represented with fewer PCs.

Table 1.1: Smallest relevant time-scale (τ_a) for a DNS case while comparing the MG-PCA and PC-transport models.

Case	DNS	MG-PCA	PC-transport
H_2	$2.32 \cdot 10^{-8}$	$2.32 \cdot 10^{-8}$	$2.55 \cdot 10^{-7}$
<i>syngas</i>	$2.34 \cdot 10^{-8}$	$2.34 \cdot 10^{-8}$	$2.60 \cdot 10^{-7}$

The third publication: "Modelling Combustion Systems with Principal Component Analysis Transport Equations", is focused on the application of the PC-transport approach in conjunction with the nonlinear regression. Various aspects of the PC-transport are analyzed, including the effect of several regression methods on accuracy. Chapter 4 gives the first demonstration of the PC-transport approach using nonlinear regression within a numerical solver. First, a simple perfectly stirred reactor is shown, and this is followed by a 2D solution within a CFD solver. The results give the first glimpse at the potential of the model.

Although the PC-transport model has several fundamental advantages, there are several remaining issues and general concerns for the model. A major advantage is seen in doing the PC-analysis on only the larger variables of the system (see Chapter 4) in an attempt to remove some of the highly nonlinear effects from the manifold. Chapter 5 discusses the potential for adding constrained equilibrium effects to the model in order to represent the species not included in the PC analysis.

The fact that the PC basis is linear can be quite troublesome seeing that the combustion systems under observation are highly nonlinear. Chapter 5 proposes a new model where nonlinear effects of the system may be included in the model by preconditioning the data. The concepts in Chapter 5 are demonstrated through *a priori* studies and show promising results.

CHAPTER 2

A NOVEL METHODOLOGY FOR CHEMICAL TIME-SCALE EVALUATION WITH DETAILED CHEMICAL REACTION KINETICS

Benjamin J. Isaac, Alessandro Parente, Chiara Galletti, Jeremy N. Thornock, Philip J. Smith, and Leonardo Tognotti

Published in Energy & Fuels, 2013, Vol. 27, p. 2255-2265

Reprinted with permission. Copyright 2013 American Chemical Society.

2.1 Abstract

Interaction between turbulent mixing and chemical kinetics is the key parameter which determines the combustion regime: only understanding such interaction may provide insight into the physics of the flame and support the choice and/or development of modelling tools. Turbulence-chemistry interaction may be evaluated through the analysis of the Damköhler number distribution, which represents the flow to chemical time-scale ratio. Large Damköhler values indicate mixing controlled flames. On the other hand, low Damköhler values corresponds to slow chemical reactions: reactants and products are quickly mixed by turbulence so the system behaves like a perfect stirred reactor. The calculation of the Damköhler number requires the definition of proper flow and chemical time-scales. For turbulent conditions, various flow time-scale can be used, such as the integral and Kolmogorov time-scales. Chemical time-scale calculation poses some issues. In literature, several examples of Damköhler calculation are reported, but in most cases, a global chemical reaction rate is used to estimate the chemical time-scale. A method for considering more complex kinetic schemes is proposed by Fox [23], who defines the chemical time-scale in terms of the inverse of the eigenvalues from the decomposition of the chemical source term Jacobian matrix. The present work aims at developing a procedure

for the calculation of the chemical time-scale (and thus of the Damköhler number) with complex kinetics starting from the analysis of the Jacobian matrix of the chemical species source terms. Emphasis is given on the dimension of the Jacobian matrix, as it is not fully understood how the species for the time-scale calculation should be chosen. In other words, one can refer to the full set of species (thus all species will have the same “weight”), but also to a subset of them. The main concept is to perform a preliminary analysis, based on Principal Variables (PV), to determine the relative importance of the chemical species, in order to select an optimal subset for the chemical time-scale calculation. The procedure is illustrated and applied for the Moderate and Intense Low-oxygen Dilution (MILD) combustion as this kind of regime shows a strong coupling between turbulence and chemistry, mainly because of slower reaction rates (due to the dilution of reactants) in comparison with conventional combustion. The methodology is further validated on a DNS data-set modelling a CO/H_2 jet flame.

2.2 Introduction

Interaction between turbulent mixing and chemical kinetics is the key aspect in combustion modelling as it determines the combustion regime. Therefore, a fundamental understanding of turbulence-chemistry interactions in reacting systems may provide the needed insight into the physics of the flame, allowing an appropriate selection or development of physical models. The Damköhler (Da) number characterizes the behavior between mixing and reaction in a system, given by the ratio of a mixing or flow time-scale to a chemical time-scale (τ_f/τ_c). The decision of the most relevant flow and chemical time-scales which control the flame structure is important in obtaining meaningful parameters which describe the system. When defining the Damköhler number for premixed flames, the flow time-scale is generally defined by the ratio of the turbulent length scale to the turbulent intensity (l/ν'), which is proportional to the integral time-scale ($\tau_I = k/\varepsilon$), being the largest turbulent time-scales in the system. The chemical time-scale, τ_c , is calculated from the ratio of the flame thickness to the laminar flame speed (δ/s_L) [90, 71], leaving a definition for the Damköhler number as:

$$Da_I = \frac{kv'}{\varepsilon l} \quad (2.1)$$

Another useful nondimensional number used in premixed flames is the Karlovits number (Ka), defined for premixed flames as the ratio of the representative chemical time-scale to the Kolmogorov mixing scale, leaving $Ka = l_F^2/\eta^2$, where η is the Kolmogorov length scale, and l_F is the flame thickness. This dimensionless number is relevant when looking at the

ability of the mixing to alter the flame physics. For premixed flames, the following general combustion regimes [68] are found:

- *Laminar Flames* ($Re < 1$) - Laminar flames with no turbulent structures effecting the physics of the system.
- *Thickened Flame* ($Re > 1, Da_I \ll 1, Ka > 100$) - A turbulent regime where the reaction time-scales of the system are much slower than the mixing time-scales. Here reactants and products are quickly mixed by turbulence as the mixing scales are small enough to enter the inner reaction layer; accordingly, the system behaves like a perfect stirred reactor leaving a system governed by the reaction scales.
- *Thin Reaction Zones* ($Re > 1, 100 > Ka > 1$) - A turbulent regime where the reaction scales of the system are larger than the smallest mixing scales, and the smallest mixing scales are not sufficiently small to enter the inner reaction layer where radicals begin to react with the fuel, but large enough to perturb the inert preheat zone, thus distorting the laminar flame structure.
- *Flamelets* ($Re > 1, Ka < 1, Da_I > 1$) - The system is governed by the mixing as the reaction times are all smaller than the smallest mixing scales; this is the case where the flame preserves a laminar flamelet shape within the smallest turbulent structures.

In contrast to premixed flames, distinct regime definition for nonpremixed combustion is difficult. The definition of a characteristic flame velocity such as that of premixed combustion is not available [69], thus complicating the calculation of a reaction time-scale. Nonpremixed flames exhibit multiple flow scales which may evolve temporally as well as have dependence on spatial coordinates, and burner flow conditions [71]; this results in multiple choices for definition of the flow scales. In addition to the Kolmogorov and integral scales discussed with premixed combustion, the diffusion layer and reaction zone are often described in nonpremixed combustion, the diffusion layer being the region where mixture fraction changes significantly and the reaction zone corresponding to a region where reaction rates are nonzero located immediately adjacent to the stoichiometric iso-surface [71]. Many authors suggest the use of the inverse of the stoichiometric scalar dissipation rate for the definition of a local mixing time τ_χ [71, 69]. The local mixing time is calculated as $\tau_\chi = 1/\chi_{st} = \left(2D \overline{|\nabla \xi''_{st}|^2}\right)^{-1}$, where ξ''_{st} is the mixture fraction fluctuation and D is the thermal diffusivity [38]. Additionally literature shows the nonpremixed Damköhler number being calculated using τ_I , and τ_η the Kolmogorov mixing time which defines the smallest

scale of momentum differences [38]. The Batchelor scale τ_B is of interest as it defines the smallest scale for scalar differences, which is comparable to the definition of the Kolmogorov scale except in terms of the reacting scalars. In general, for gaseous mixtures, one assumes, $\tau_\eta \approx \tau_B$. Three definitions of the Damköhler number for nonpremixed combustion can be calculated depending on the choice of the mixing time: the integral Damköhler number (Da_I), the local mixing Damköhler number (Da_χ), which uses the local mixing time τ_χ , and the Kolmogorov Damköhler number (Da_η). Several authors have attempted to characterize nonpremixed combustion regimes. An example of one was demonstrated by Poinot [71], where Da and the turbulent Reynolds (Re_t) number, defined as $\frac{\tau_I}{\tau_\eta} = \frac{k/\varepsilon}{(\nu/\varepsilon)^{1/2}} = \sqrt{\left(\frac{k^2}{\nu\varepsilon}\right)} = \sqrt{Re_t}$, are used to characterize the regimes. When the Damköhler number is large enough, the laminar flamelet assumption (LFA) applies, leaving a transitional Damköhler number (Da_I^{LFA}) where the combustion regime changes. Given a sufficiently small Damköhler number, extinction occurs; here the transitional Damköhler number Da_I^{EXT} is defined. The following nonpremixed combustion regimes are generally described in terms of the Damköhler number.

- *Laminar* ($Re_t < 1$) - Simple diffusion flames with no turbulent structures effecting the diffusion of fuel and oxidizers.
- *Flamelet* ($Re_t > 1, Da_I \geq Da_I^{LFA}$) - A turbulent regime where mixing scales are larger than reaction scales thus preserving the steady laminar flamelet.
- *Unsteady* ($Re_t > 1, Da_I^{LFA} > Da_I > Da_I^{EXT}$) - A regime where mixing scales produce instability in the flame front.

The determination of chemical time-scales for turbulent combustion systems is particularly difficult as detailed reaction mechanisms are often required for adequate description of the combustion process. A definition of the laminar flame velocity s_L , used for the calculation of $\tau_c (\delta/s_L)$, does not exist for nonpremixed combustion. The chemical time (τ_c) is determined from several different methods, including activation energy asymptotics[41], global chemistry assumption [8], or the critical scalar dissipation rate at quenching $\tau_q = 1/\chi_q$, has been used for estimation of τ_c with complex chemistry cases [38]. At this time, a clear definition for τ_c for complex chemistry systems involving detailed kinetic mechanisms is needed, and it is the focus of this work. In most cases, a global chemistry assumption is made to simplify the estimation of τ_c . An example of this is given by Kuo [40] where the following definition is used:

$$Da_I = \left(\frac{\nu K_r^2}{\varepsilon} \right) \quad (2.2)$$

where ν is the kinematic viscosity, ε is the dissipation of turbulent kinetic energy, and K_r is the kinetic constant of the global reaction. On the other hand, Fox [23] provides a method for considering more complex kinetic schemes, suggesting that the chemical time-scale can be defined in terms of the eigenvalues of the $Q \times Q$ Jacobian matrix \mathbf{J} of the chemical source terms, whose elements J_{ij} are given by (for an isothermal case):

$$J_{ij} = \frac{\partial R_i}{\partial Y_j} \quad (2.3)$$

The eigenvalue decomposition of \mathbf{J} leaves an expression for chemical time-scales as:

$$\tau_a = \frac{1}{|\lambda_a|} \quad (2.4)$$

where λ_a is the set of eigenvalues from $a = 1, \dots, Q$. Here Q is the number of species in the detailed mechanism. In a complex kinetic scheme, for which the time-scales can range over several orders of magnitude, the slowest chemical time-scale should be chosen for the estimation of the Damköhler number:

$$\tau_c = \max(\tau_a) \quad (2.5)$$

Such an approach was recently applied by Rehm et al. [78] who calculated the Damköhler number for a gasification system using the five most abundant species to define the Jacobian matrix. Retaining all the species of the kinetic mechanism may lead to the determination of nonmeaningful time-scales, due to the complete inactivity of some species in specific regions of the flame. The choice of the species to be retained is not to date established and is generally made by looking at the major species, as done in the work by Rehm et al. In the present work, the selection of species to be retained is addressed and a newly proposed method is presented in the subsequent section.

Upon determination of an appropriate expression for τ_c and τ_f , the evaluation of the Damköhler number can easily allow one to identify the predominant combustion regime and choose the appropriate turbulent combustion model. Existing turbulent combustion models are generally well suited for (1) high Damköhler numbers where mixing dominates the process or (2) low Damköhler numbers where chemistry dominates the physics and finite rate chemistry models are required. An example of a high Damköhler number model is the Steady Laminar Flamelets Model (SLFM) [67, 65], which treats the flame as an ensemble of steady laminar diffusion flames. Here the mixture fraction variable is used as an independent variable which defines the thermochemical state-space; mixture fraction fluctuations are generally specified using an assumed shape of the PDF of the mixture

fraction. The shape of the PDF is characterized by transporting the mean mixture fraction and the variance of the mixture fraction. In lower Damköhler flows, turbulent structures can enter the flame preheating zone and further mix and distort the flame front; these unsteady effects require a modelling approach with higher coupling between the chemical reactions and the turbulent mixing. A model such as Eddy Dissipation Concept (EDC) [43, 44, 45, 46] transports the species involved in a detailed reaction mechanism, and treats the flame as an ensemble of perfectly stirred reactors (PSR) where the PSR residence time is a function of the local mixing time-scales. This allows for a more complex chemistry tracking approach, while coupling the turbulent structures to the chemistry physics.

A particular combustion regime of interest in terms of the Damköhler analysis is the flameless (or MILD) combustion regime. This regime is characterized by a strong coupling between turbulence and chemistry, because of slower reaction rates (due to the dilution of reactants) with respect to conventional combustion [8]. It is a widespread opinion that for such a combustion regime, the Damköhler number approaches unity [25]. Indeed many modelling investigations have shown that high Damköhler number approaches such as SLFM are not suited for this combustion regime due to the slow chemistry [12]. Encouraging results have been obtained through the EDC model [46, 10, 26, 59, 3, 58], especially for its capability of handling detailed kinetic schemes [60]. However, some discrepancies are still observed when using EDC and model modifications have been proposed in the literature for better capturing flameless conditions [1, 4].

The objective of the present paper is that of defining a methodology for the determination of the principal variables of a reacting system, to allow the determination of a chemical time-scale τ_c based on complex reaction schemes, so that a meaningful Damköhler number may be calculated. Such a formulation is interesting for combustion processes where detailed kinetics need to be taken into account like in flameless combustion. The proposed choice of the size of the Jacobian matrix and of the variables that should be involved in the chemical time-scale calculation becomes fundamental. The present paper proposes a methodology based on principal variable analysis for the selection of the variables carrying most of the relevant information. In the following section, the methodology is first presented with an introduction to principal component analysis (PCA) and principal variables. Then a discussion on the size of the Jacobian follows, to rigorously determine the minimum number of species which should be included in the Jacobian calculation. Finally, results are presented for a flameless combustion data-set [19], and as a consistency check, the approach is demonstrated on a DNS data-set for nonpremixed syngas combustion.

2.3 Methodology

2.3.1 Principal component analysis

PCA [33, 31] is a statistical technique employed in the analysis of multivariate data-sets, for detecting the directions that carry most of the data variability, thus providing an optimal low-dimensional manifold of the system. For a data-set, \mathbf{X} , consisting of n observations of Q variables, the sample covariance matrix, \mathbf{S} , of \mathbf{X} can be defined as $\mathbf{S} = 1/(n-1) \mathbf{X}^T \mathbf{X}$. Recalling the eigenvector decomposition of a symmetric, nonsingular matrix, \mathbf{S} can be decomposed as $\mathbf{S} = \mathbf{A} \mathbf{L} \mathbf{A}^T$, where \mathbf{A} is the $(Q \times Q)$ matrix whose columns are the eigenvectors of \mathbf{S} , and \mathbf{L} is a $(Q \times Q)$ diagonal matrix containing the eigenvalues of \mathbf{S} in descending order, $l_1 > l_2 > \dots > l_p$. The covariance matrix indicates the level of correlation between the nondimensional variables. Values close to zero denote uncorrelated variables whereas correlations close to one indicate strongly correlated variables. Based on the correlation values, the redundant, less important information contained in the original data-sets can be easily removed. Once the decomposition of the covariance matrix is performed, the Principal Components (PCs), \mathbf{Z} , are defined by the projection of the original data onto the eigenvectors, \mathbf{A} , of the covariance matrix, \mathbf{S} , $\mathbf{Z} = \mathbf{X} \mathbf{A}$. Then, the original variables can be stated as a function of the PC as $\mathbf{X} = \mathbf{Z} \mathbf{A}^T$, being \mathbf{A} orthonormal and, hence, $\mathbf{A}^{-1} = \mathbf{A}^T$. Nevertheless, the main objective of PCA is to replace the p elements of \mathbf{X} with a much smaller number, q , of PC, preserving at the same time the amount of information originally contained in the data. If a subset of size $q \ll Q$ is used, the truncated subset of PC is $\mathbf{Z}_q = \mathbf{X} \mathbf{A}_q$. This relation can be inverted to obtain:

$$\mathbf{X}_q = \mathbf{Z}_q \mathbf{A}_q' \quad (2.6)$$

where \mathbf{A}_q is the matrix obtained by retaining only the first q columns of \mathbf{A} . The linear transformation provided by Equation 2.6 ensures that the new coordinate axes identified by PCA are orthogonal and they provide independent and decreasing contributions to the amount of original variance explained by the PC. Thus, if only the subset \mathbf{A}_q of \mathbf{A} is retained, \mathbf{X}_q represents the best q -dimensional approximation of \mathbf{X} in terms of squared prediction error. Variables are generally centered before the PCA analysis, to convert observations to fluctuations on the mean; moreover, scaling is applied when the elements of \mathbf{X} have different units or when they have different variances, as is the case for this investigation. A centered and scaled variable can be defined as [61, 63]:

$$x_j = \frac{x_j - \bar{x}_j}{d_j}. \quad (2.7)$$

where d_j is the scaling parameter adopted for variable x_j . Several scaling options are available, including normalization by the variable range, standard deviation, maximum, and average values [63, 5]. The present paper uses the standard deviation as scaling parameter. This ensures that all the elements of the scaled \mathbf{X} matrix have a standard deviation equal to one and a covariance ranging from zero to one giving them similar relevance.

In the case of averaged data-sets, the insensitivity of the eigenvector matrix \mathbf{A} to filtering was recently demonstrated by Biglari and Sutherland [5]. Their results suggest the structure of PCA defined by \mathbf{A} remains relatively unchanged while using several filter widths, meaning the low-dimensional manifold for the state variables is relatively insensitive to filtering. This is significant as an averaged data-set may be sufficient for extracting PVs.

The low-dimensional manifold discussed in the context of PCA is fundamentally different than low-dimensional manifolds used in many other models, such as the thermodynamic manifold of rate-control constrained equilibrium (RCCE) [83], or the reaction-diffusion manifold of flame prolongation of ILDM (FPI) [21, 22]. Here the manifold is empirically based on the state-space from computed turbulent combustion data, whereas other techniques construct the manifold from the solution of the conservation equations for simplified systems such as laminar flames [72].

2.3.2 Principal variables

PVs represent an attempt to gain a physical understanding from principal component analysis. PV algorithms try to link the PC back to a subset of the original variables, which satisfies one or more optimal properties of PCA, such as the maximization of the variance of the original data \mathbf{X} . Then one can partition the set of variables into $\mathbf{X}(1)$ and $\mathbf{X}(2)$. Where the expression for the sample covariance matrix is:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \quad (2.8)$$

The partial covariance for $\mathbf{X}(2)$ given $\mathbf{X}(1)$ can be expressed as:

$$\mathbf{S}_{22,1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}. \quad (2.9)$$

With respect to Equation 2.9, PV techniques attempt to minimize the information carried by the covariance matrix $\mathbf{S}_{22,1}$, by considering only the most important variables within the data-set. Several PV methods exist in the literature [33]. In the present work, the B2 backward method is considered. A PC analysis is performed on the original matrix of Q variables and n observations. The eigenvalues of the covariance/correlation matrix

are then computed and a criterion is chosen to retain q principal variables. Starting with the last eigenvector of \mathbf{A} corresponding to the smallest eigenvalue, the variable associated with the highest eigenvector coefficient is discarded as the latter is highly correlated with a component carrying the least variance. The process is then repeated for the next highest eigenvalue until q PVs remain, leaving $S_{22,1}$ as a $(Q - q) \times (Q - q)$ dimensional matrix. The trace of $S_{22,1}$ is employed to determine the number of variables q to be retained, once a value for the loss in variance γ is set:

$$\frac{\text{trace}(S_{22,1}(q))}{\text{trace}(S)} \lesssim \gamma \quad \text{for } q = 1, 2, \dots, Q \quad (2.10)$$

The term $\frac{\text{trace}(S_{22,1}(q))}{\text{trace}(S)}$ can be interpreted as the lost variance of \mathbf{X} by selecting the subset q . Generally, a value of $\gamma = 0.01$ is chosen, which implies a retention of 99% of the variance in the system.

2.3.3 Jacobian matrix down-sizing

The variables extracted with the principal variable algorithms are used to compute a subset of the full Jacobian matrix, only including the derivatives with respect to the selected principal variables. This allows the determination of the modes to be compared with the ones provided by the full Jacobian matrix, including all the species involved in the original detailed kinetic mechanism. In all cases, the determination of the Jacobian matrix is performed with an *in house* MATLAB® code JACOBEN. The code is particularly interesting as it formulates the chemical source term equations as symbolic expressions and then uses the symbolic differentiation function in MATLAB® to form the analytical expressions for the derivatives of the chemical source terms with respect to chemical species. The code requires the chemical mechanism to be in CHEMKIN format, as well as all thermodynamic state-space parameters describing the turbulent combustion, including the species mass fractions and temperature. A Jacobian matrix is then evaluated for every observation provided in the thermodynamic state space input file.

Figure 2.1 summarizes the process used in order to calculate the Damköhler number. Principal variables analysis as well as the mixing time-scales are calculated directly from simulation data. The chemical mechanisms provide the chemical reactions which are used to form the symbolic expressions for the chemical source terms, as well as the thermodynamic information needed to calculate the equilibrium constants. Upon forming the Jacobian matrix, the PV analysis identifies the subset of the Jacobian matrix used to calculate the chemical time-scales.

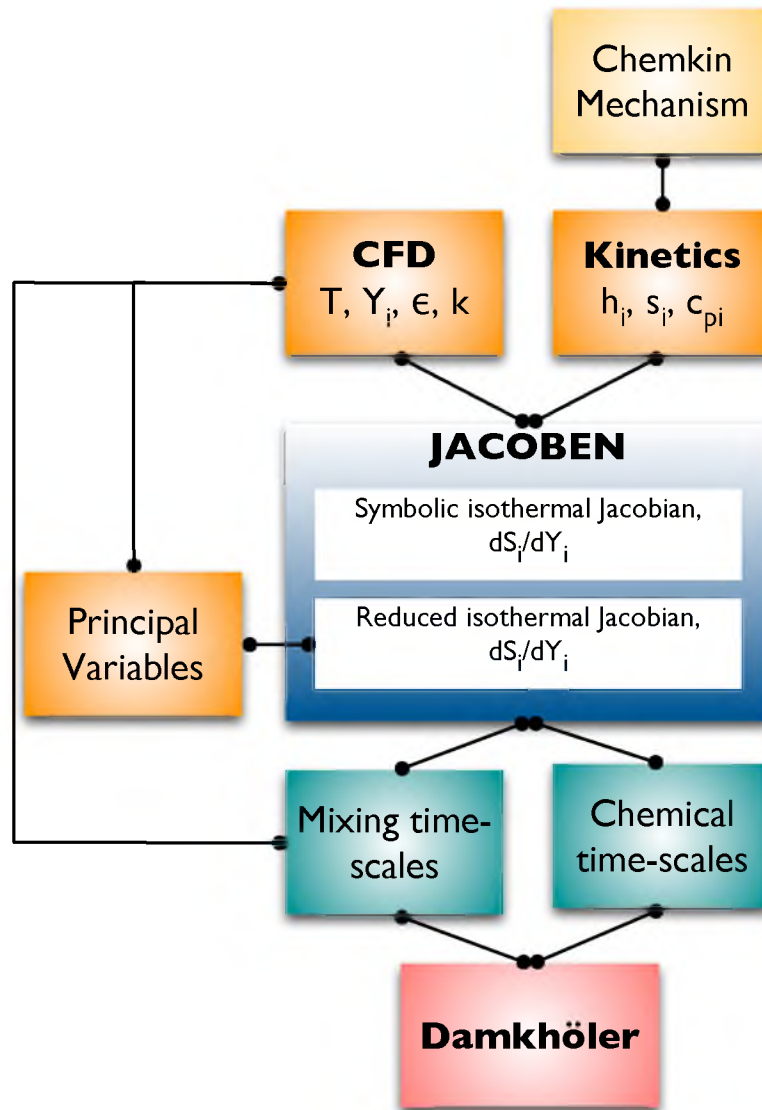


Figure 2.1: Flow diagram describing the process of analysis used to obtain the Damköhler number. Here the subscripts i , r , and pv denote species, reaction number, and principal variable, respectively.

2.4 Test cases

First, a demonstration of the chemical time-scale approach is given based on a simulated flameless combustion data-set where a reduced Damköhler number is expected. The simulation data were provided by Aminian et al. [4] in reference to the jet in a hot co-flow (JHC) burner designed by Dally [19] to emulate flameless combustion conditions. The experiments by Dally consist of a jet with a CH_4/H_2 mixture (inner diameter of 4.25 mm) within an annulus co-flow (inner diameter of 8.2 mm) of hot oxidizer gases from a porous bed burner mounted upstream of the exit plane. The entire burner is placed inside a wind tunnel feeding air at the same velocity as the hot co-flow. The experiments operate with a jet Reynolds number of around 10,000 and different oxygen mass fraction, i.e. 3% (HM1), 6% (HM2) and 9% (HM3) in the co-flow. The available experimental data consist of 280,000 measurements of temperature and concentration of major (CH_4 , H_2 , H_2O , CO_2 , N_2 , and O_2) and minor species (NO , CO , and OH) at different locations. A detailed description of the systems and tests can be found in the work by Dally [19].

The JHC was modeled with the Fluent 6.3 software by Ansys Inc. A 2D axisymmetric domain was chosen with a structured grid containing 25,000 cells. The steady-state Reynolds-Averaged Navier-Stokes (RANS) equations were solved with a modified version of the $k-\epsilon$ turbulence model (i.e. imposing $C_{\epsilon 1} = 1.6$ for round jets) [51, 3]. The KEE-58 mechanism [7] consisting of $Q = 17$ species and 58 reversible reactions was used. Turbulence-chemistry interactions were modeled with the EDC model. The constant of the residence time in the fine structures was set to 1.5 as this was found to improve substantially the predictions of temperature and chemical species in flameless conditions [4]. Although the reactions rates for the species in this model are directly tied to the mixing in the system, the model parameters have been trained sufficiently to ensure the major species profiles are accurate when compared with the experimental results. For the boundary conditions, velocity inlet conditions were given to the unmixed fuel jet, co-flow oxidizer, and tunnel air, paying particular attention to turbulent intensity [3]. The simulation results obtained from the simulation of the JHC burner have been successfully validated [4] against the available experimental data and they represent an ideal data-set for testing the proposed methodology for chemical time-scale calculation in case of complex kinetic mechanisms. Figure 2.2 shows the simulation results of temperature T for cases HM1, HM2, and HM3 as a function of the mean mixture fraction. For the demonstration of the chemical time-scale approach, the simulation data-set consisted of $n = 25,000$ points, with $Q = 17$ variables, being the 17 fine structure mass fractions of the chemical species.

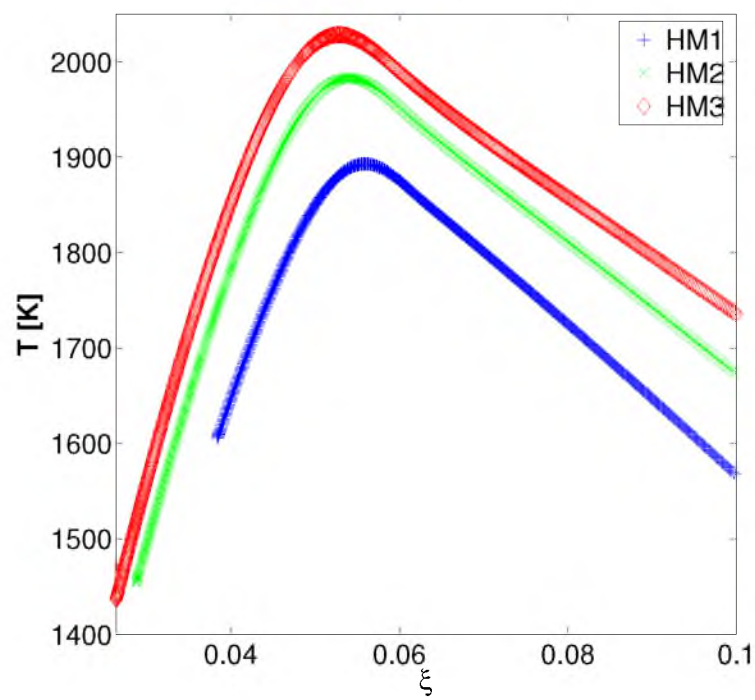


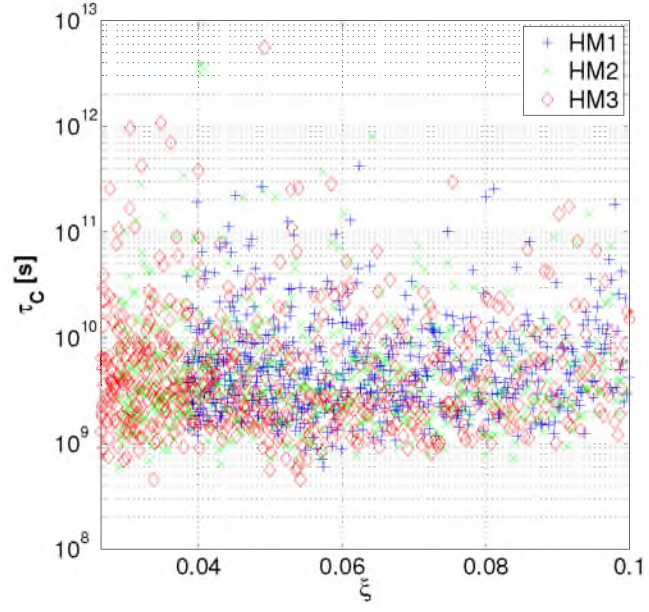
Figure 2.2: Temperature as a function of mixture fraction ξ for cases HM1, HM2, and HM3.

Second, in order to demonstrate consistency for the chemical time-scale approach, an additional analysis on Direct Numerical Simulation (DNS) data, created by Sutherland [86] using Sandia National Laboratories S3D [9] code, is shown. The code uses 8th order explicit finite difference derivatives, and a 4th order, six-stage explicit Runge-Kutta time integrator. Here a case with $Re = 4500$, an inlet fuel stream containing 50% CO , 10% H_2 , and 40% N_2 (by volume), and an oxidizer stream containing air are simulated. The domain consists of a 2D rectangular mesh containing 2160 by 720 grid points evenly spaced. The oxidation of the CO/H_2 mixture was described using the mechanism developed by Yetter et al. [92]. The mechanism consists of $Q = 12$ species: H , O_2 , O , OH , H_2 , H_2O , CO , CO_2 , HO_2 , H_2O_2 , HCO , N_2 and 33 chemical reactions with the simulation data-set containing $n = 4.7$ million points taken from several snapshots in time with even intervals of 0.5 seconds.

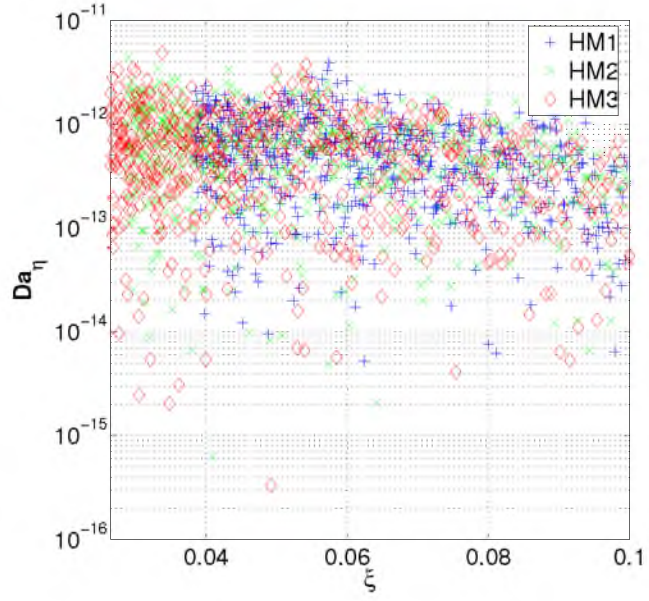
2.5 Results and discussion

We first analyze the flameless combustion data-set. Figure 2.3a shows the largest chemical time-scale of the system (τ_c) using Equation 2.5, and Figure 2.3b the Damköhler number Da_η , all as a function of the mean mixture fraction ξ . Here ξ is calculated using Bilger's mixture fraction formula [6] from the species, as in Dally and Christo [19, 10]. The plotted data come from the axis of the cylindrical burner for all three systems (HM1-HM3). In Figure 2.3, the results for τ_c and the Damköhler number are calculated using the full Jacobian matrix for the chemical time-scales (Equations 2.3-2.5) and the Kolmogorov mixing scale. From the analysis, it is clear that keeping all of the variables in the Jacobian matrix does not help in identifying the relevant processes for the system under investigation. In particular, if no filtering is applied to the original thermo-chemical state variables, the analysis will point out the existence of the extremely slow time-scales of the nonreacting species, as shown in Figure 2.3, leading to Da_η values close to zero for all three systems.

A first attempt is then that of filtering out from the Jacobian analysis all the slow scales, i.e. all values above 1000 seconds (limit for slow chemistry processes according to Fox [23]). This results in the plots of Figure 2.4, showing the time-scales (Figure 2.4a) and Da_η values (Figure 2.4b) for a filtered Jacobian matrix. However, the new time-scale analysis does not provide a clear insight into the investigated combustion system. The Da_η values obtained for the three cases appear similar in magnitude and simply shifted along the mixture fraction axis going from case HM1 to case HM3. More importantly, the Da_η values are in all cases fairly large (≈ 8) near the stoichiometric mixture fraction, thus far from what would be expected in flameless conditions. The results shown in Figure 2.3 indicate that keeping all

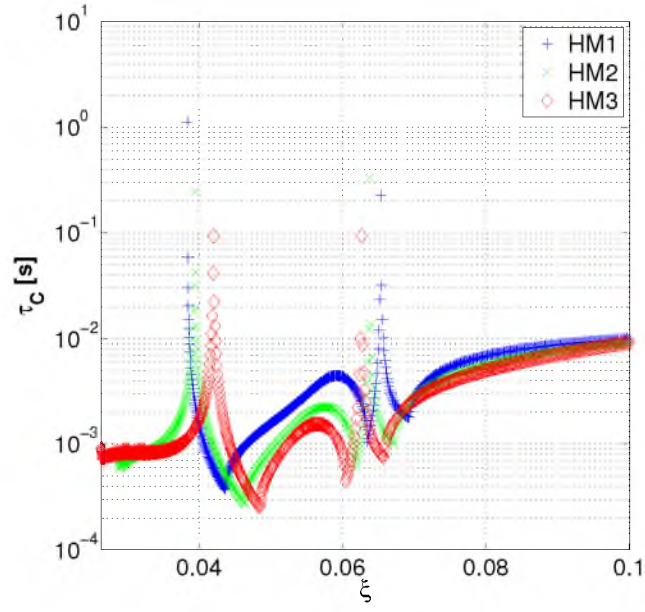


(a)

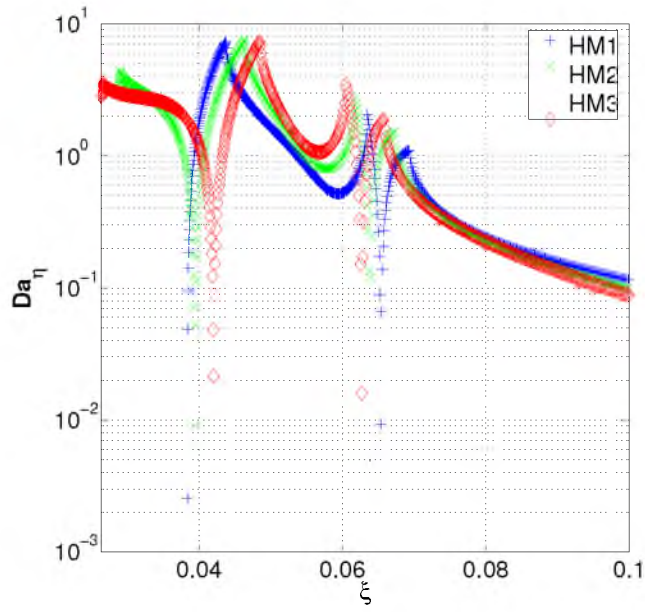


(b)

Figure 2.3: Chemical time-scale τ_c (a), and Damköhler values Da_η (b) as a function of mixture fraction ξ . Full Jacobian matrix and unfiltered time-scales.



(a)

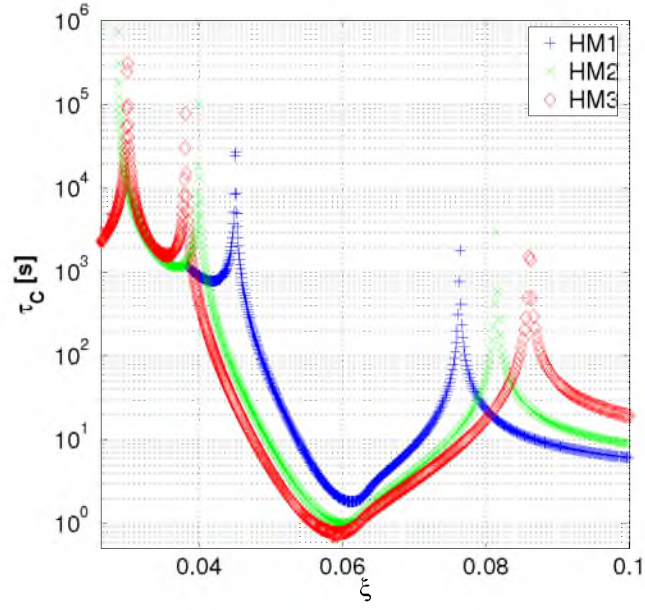


(b)

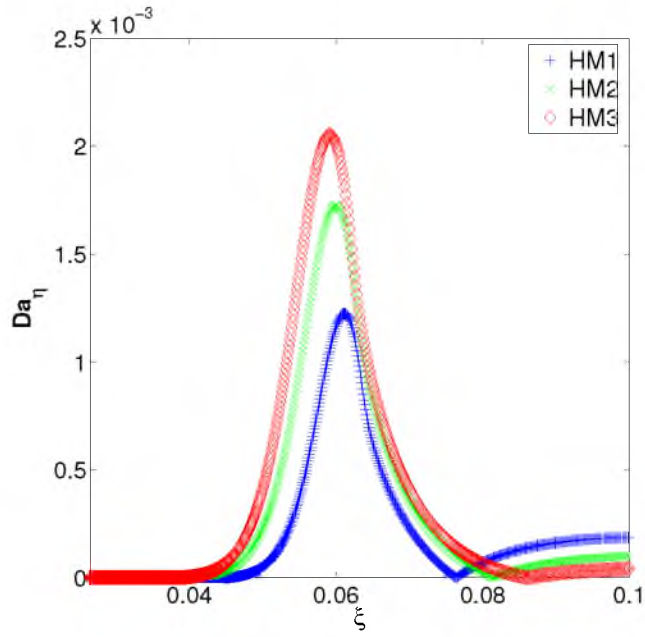
Figure 2.4: Chemical time-scale τ_c (a), and Damköhler values Da_η (b) as a function of mixture fraction ξ for the full Jacobian matrix. Time-scales above 1000 seconds have been removed.

the thermo-chemical state variables in the Jacobian matrix does not allow identifying the controlling chemical time-scale of the system. It is therefore very important to identify the relevant variables for the time-scale analysis through a rigorous selection method.

According to the method shown by Rehm [78], one could now select the major species in the system in order to capture the principal τ_c of the system. Selecting a mean mass fraction criteria ≥ 0.01 leaves the major species: CH_4 , H_2 , O_2 , CO_2 , H_2O , and CO . Figure 2.5 shows the resultant τ_c and Da_η given the major species as PVs. It is observed that the evaluation of Equations 2.3, 2.4, and 2.5 with the Jacobian matrix being calculated with only the major species alone does not guarantee that the most meaningful or relevant chemical time-scale will be expressed; in this case, it leaves a large maximum time-scale, due to the fact that one or more of the species may contain dormant reaction rates (in this case, the addition of CO is highly sensitive to the large time-scale), thus yielding an unrealistic Damköhler number. The principal variables of the system which lead to the calculation of a realistic τ_c can be identified using the methodology shown above. Equation 3.18 provides the normalized trace of the lost covariance given a guess for q . The criterion of Equation 3.18 is met when $q \geq 10$ for all three cases. Figure 2.6 shows the normalized trace of the variance which is lost based on a given value of q . It can be inferred that retaining ten PVs will yield a 1% loss of variance explained by keeping all of the variables. Figure 2.7 shows in black and red, respectively, the chemical time-scales associated to the Jacobian matrix of system HM1 with all the state variables and with ten PVs, determined with method B2 (see Methodology Section). It can be observed that the slowest chemical time-scale, which is mostly pertaining to the species CO_2 in the present case, allows for a description of the slow governing dynamics of the reacting system without showing the peaks displayed by the largest (meaningful) time-scale of the full system, which come from the complex interactions between the different chemical species. Figure 2.8 shows the chemical time-scale, τ_c , and Damköhler, Da_η , values as a function of mixture fraction, ξ , for the Jacobian matrix calculated using ten principal variables for all three flames. Differently from Figure 2.3, the results indicate a meaningful trend, showing increasing Da_η values when going from HM1 to HM3, i.e. increasing the oxygen in the co-flow from 3% to 9% and moving towards conventional flame conditions. It is interesting to note that the evaluation of τ_c remains constant while using one to eleven PVs, as the first PV identified by the system is in fact CO_2 . Upon addition of the twelfth PV (CO) the analysis of the time-scales shows again the appearance of large chemical times, which are related to inactive thermo-chemical variables and which should not be considered in the analysis. This



(a)



(b)

Figure 2.5: Chemical time-scale τ_c (a), and Damköhler values Da_η (b) as a function of mixture fraction, ξ , using only the major species in the reduced Jacobian matrix.

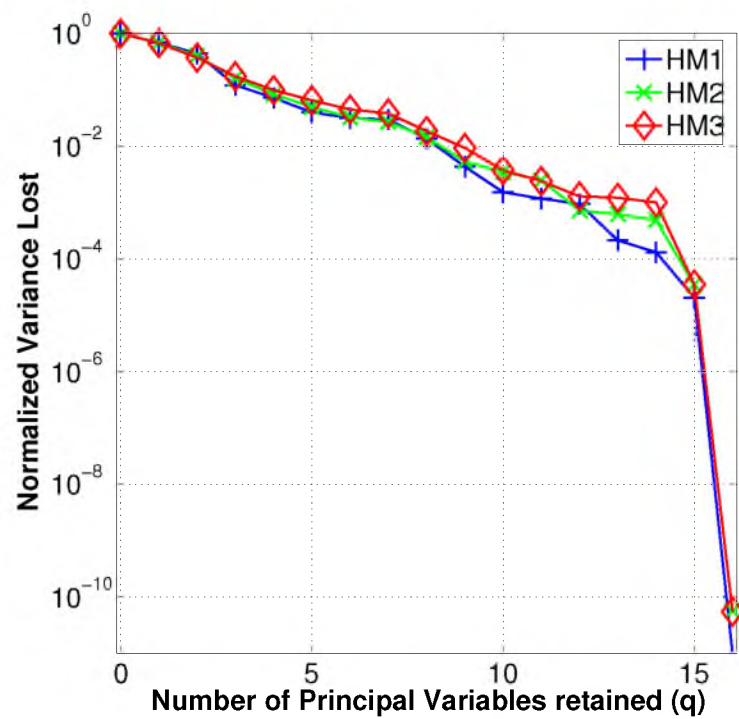


Figure 2.6: Trace plot for cases HM1, HM2, and HM3. Y axis gives a normalized variance which is lost based on the selection of q , the x-axis.

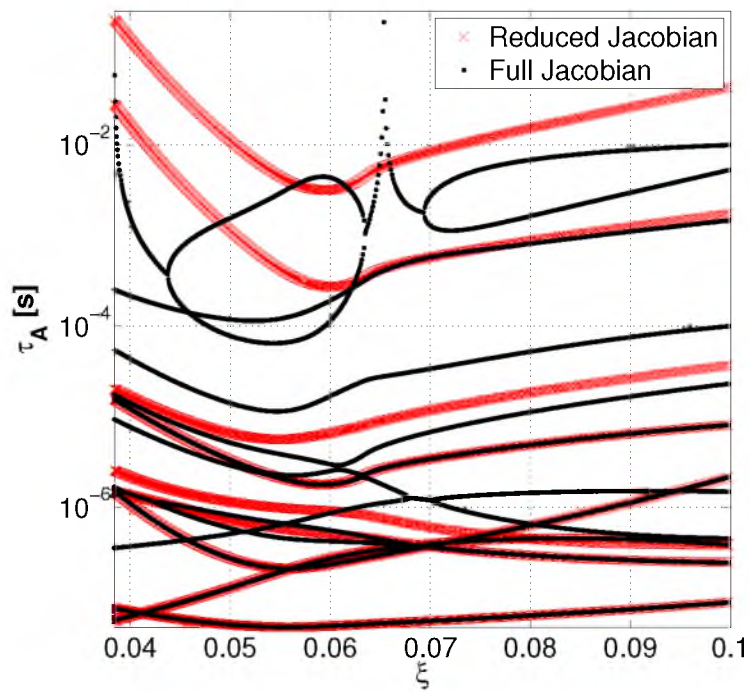
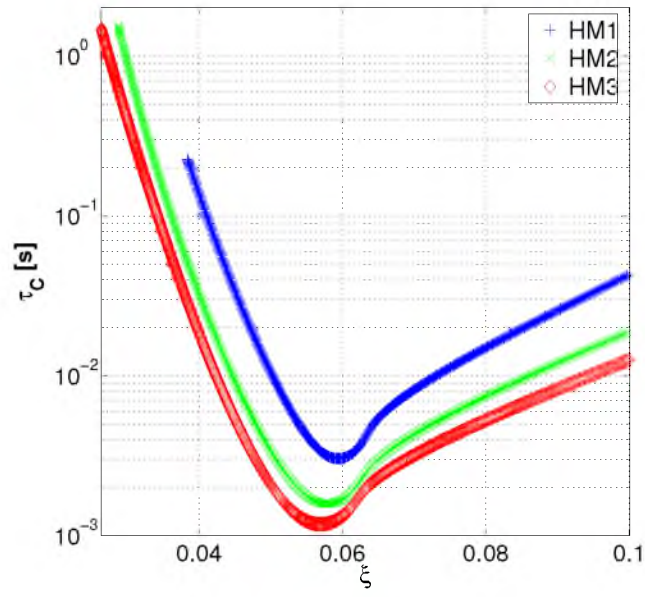
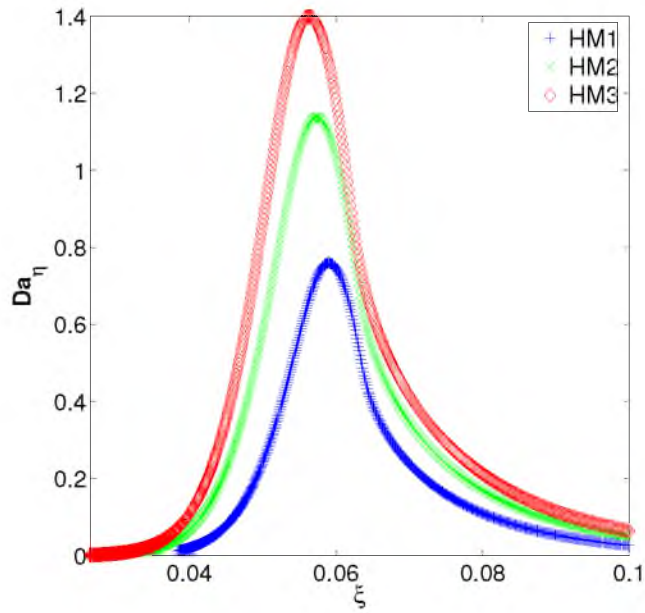


Figure 2.7: Chemical time-scales associated to the full (black) and reduced (red) Jacobian matrix (red line). Reduced Jacobian obtained using 10 PVs, determined using method B2. HM1 case.



(a)



(b)

Figure 2.8: Chemical time-scale τ_c (a), and Damköhler values Da_η (b) as a function of mixture fraction, ξ . Jacobian matrix restricted to ten principal variables.

is confirmed by the analysis of Figure 2.9 , which shows the Da_η values as a function of ξ from data taken along the burner axis. When the number of PVs is greater than eleven, the largest Da_η values drop to zero (red markers) due to the appearance of large chemical time-scales. The proposed methodology provides a very robust way for the determination of the limiting time-scale associated to a chemically reacting system. In particular, it can provide the variables that should not be included in a time-scale analysis as they do not add useful information being inactive species.

Examination of the Damköhler number using different mixing scales is shown in Figures 2.10, 2.11, and 2.12. Here the area between the Damköhler numbers calculated using the integral time-scale Da_I and Kolmogorov time-scale Da_η is shown. This Damköhler number space is significant as it shows the range of the Damköhler numbers accessible between the largest and smallest mixing scales, and can help in giving an idea of appropriate modelling strategies. Figure 2.10 shows that the Kolmogorov mixing scales for the HM1 case near ξ_{st} are smaller than the limiting chemical time-scales $Da_\eta < 1$. In this case, the smallest turbulent eddies are able to mix at a faster rate than the limiting reaction time. It is observed that $Da_I > 1$ for all three cases, meaning reaction occurs faster than the mixing by the largest turbulent motions in the system. Because the integral time-scales are much larger than the reaction scales, the system depends on the incorporation of mixing physics. As would be expected if $Da_I < 1$, a simple perfectly stirred reactor (PSR) model could accurately represent the entire system. Dilution of O_2 leads to slower limiting reactions and lower Damköhler numbers, which brings the system closer to a PSR. Figures 2.11 and 2.12 show the smallest mixing scales are larger than the reaction scales; thus, the flame should preserve more of a flamelet like shape and within the reaction zone, temperature gradients are higher. Figure 2.13 shows the contour plots of Da_η for the three cases including stream lines of constant temperature. As one would expect, a higher O_2 content in the co-flow stream leads to faster combustion, and higher Damköhler numbers. The effect of this is seen in Figure 2.13. In the case of diluted O_2 (HM1), the reacting gases penetrate further along the axis of the burner, leaving lower temperatures, Damköhler numbers, and higher chemical time-scales. The chemical time-scale approach by Fox [23] allows additional insight into the reaction times by observing weights in the eigenvector matrix from the decomposition of the Jacobian of the source terms. Figure 2.14 shows the weights for the eigenvectors corresponding to the calculated time-scales. The stacked bar chart shows the sensitivity of the chemical time mode to the principal variables. As would be expected the slow time-scales correspond to species such as CO_2 , O_2 , N_2 and rapid time-scales

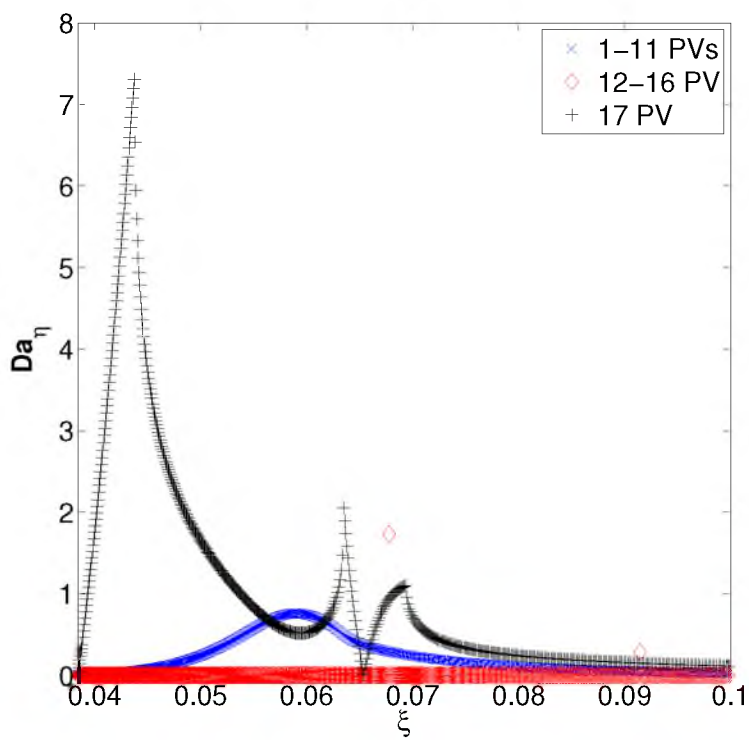


Figure 2.9: Da_η values as a function of the mixture fraction along the burner axis for flame HM1, varying the number of principal variables.

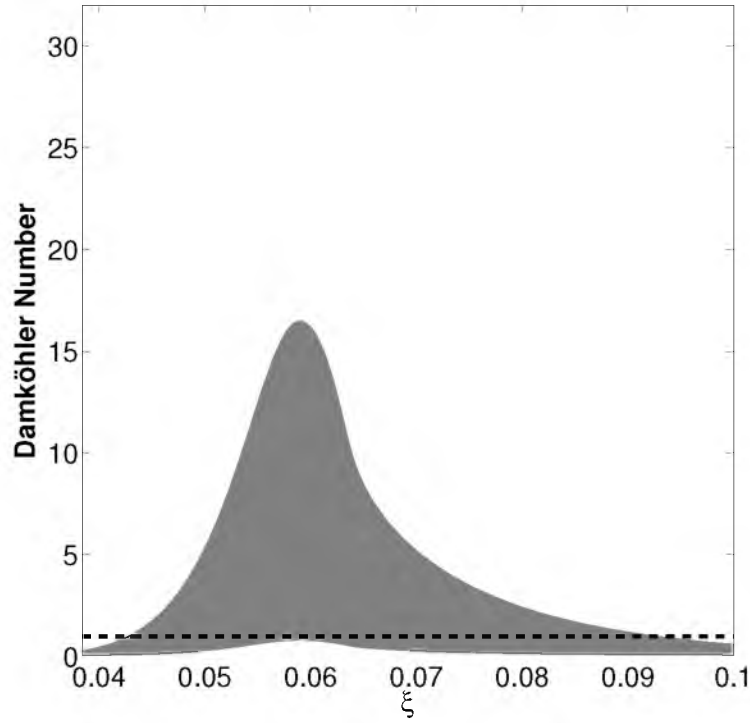


Figure 2.10: Damköhler numbers calculated using the integral and the Kolmogorov time-scales. The area bounded between the curves represents the range of Damköhler numbers between the largest mixing scales (Da_I) and the smallest mixing scales (Da_η). The dashed line represents a Damköhler number of one along the entire mixture fraction space. HM1 data-set.

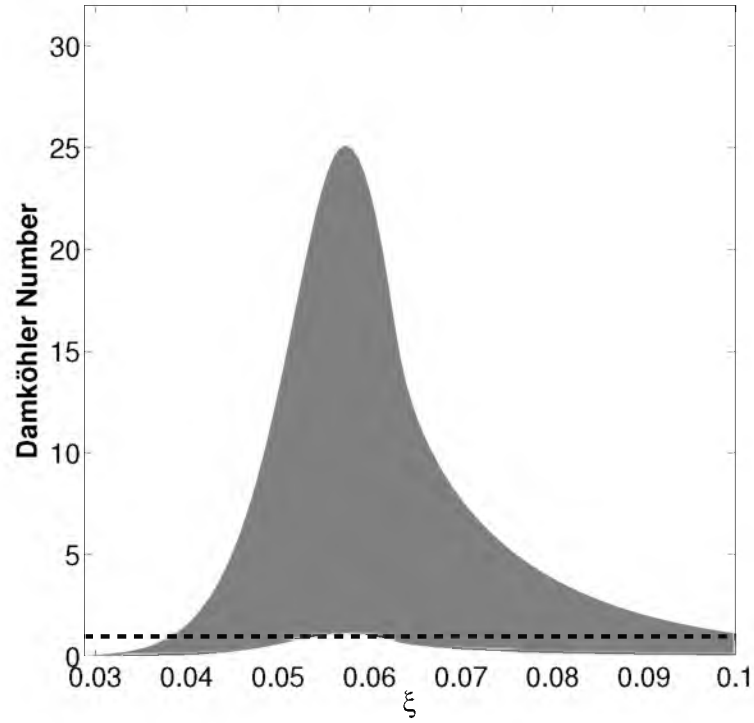


Figure 2.11: Damköhler numbers calculated using the integral and the Kolmogorov time-scales. The area bounded between the curves represents the range of Damköhler numbers between the largest mixing scales (Da_I) and the smallest mixing scales (Da_η). The dashed line represents a Damköhler number of one along the entire mixture fraction space. HM2 data-set.

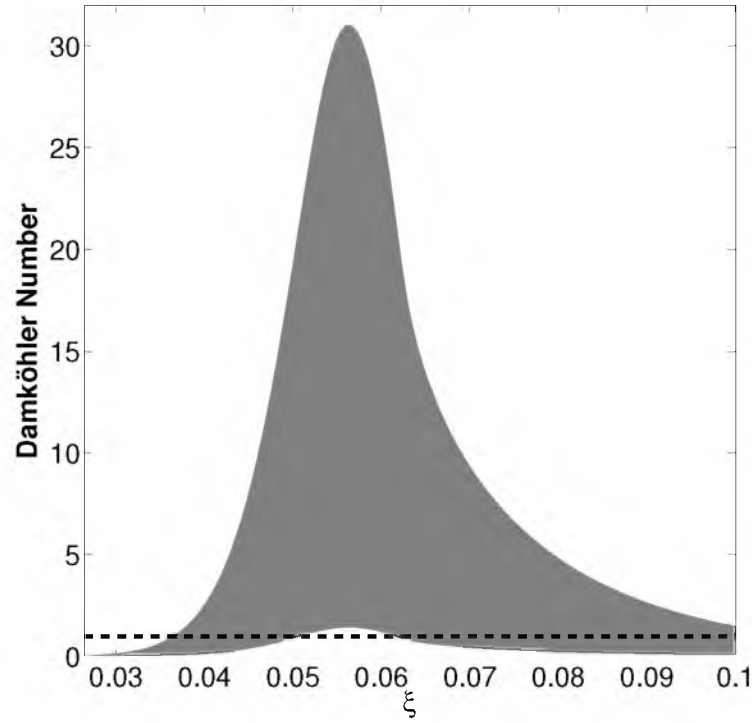


Figure 2.12: Damköhler numbers calculated using the integral and the Kolmogorov time-scales. The area bounded between the curves represents the range of Damköhler numbers between the largest mixing scales (Da_I) and the smallest mixing scales (Da_η). The dashed line represents a Damköhler number of one along the entire mixture fraction space. HM3 data-set.

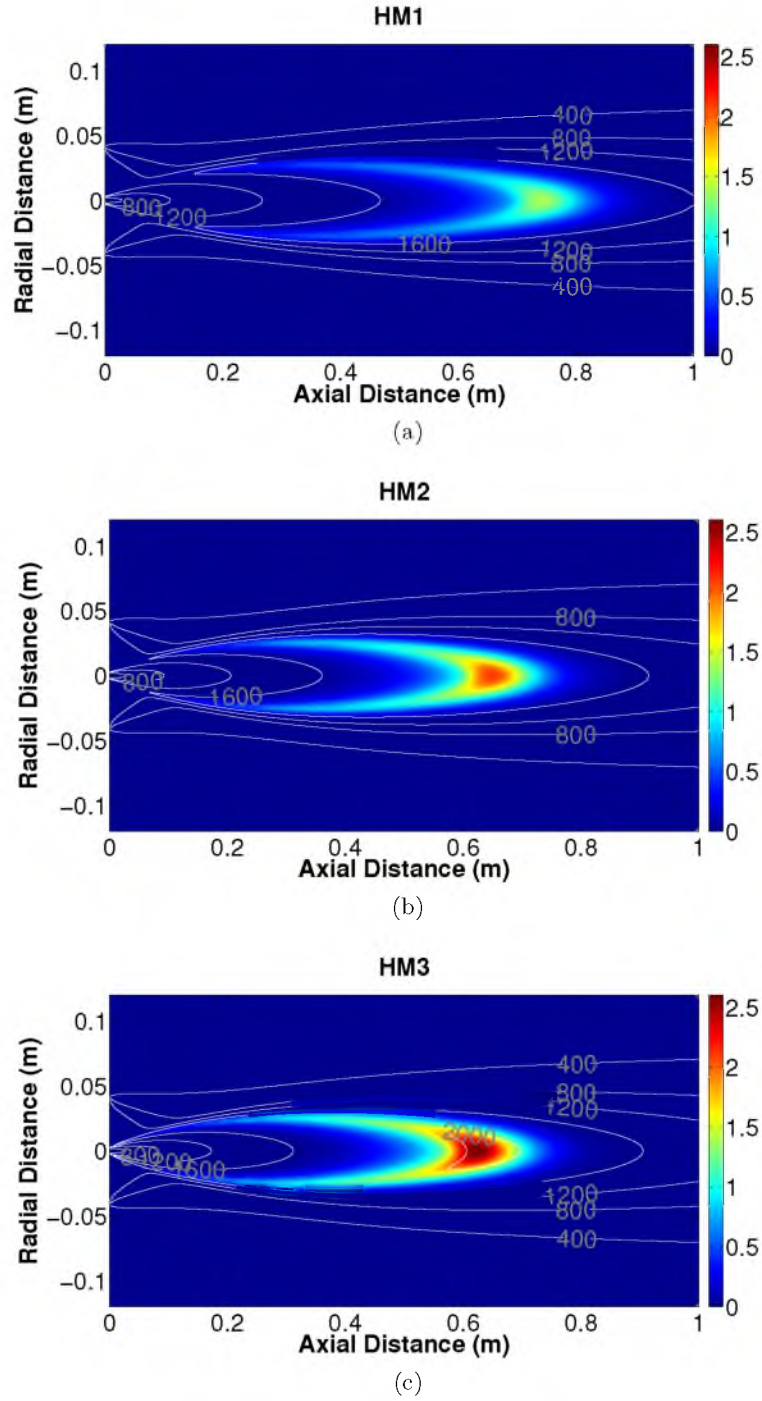


Figure 2.13: Contours showing Da_η values along the radial and axial direction of the burner. Colors in the graphs represent Da_η and isolines show the temperature. Plots (a), (b), and (c) show the contours for the cases HM1, HM2, and HM3, respectively.

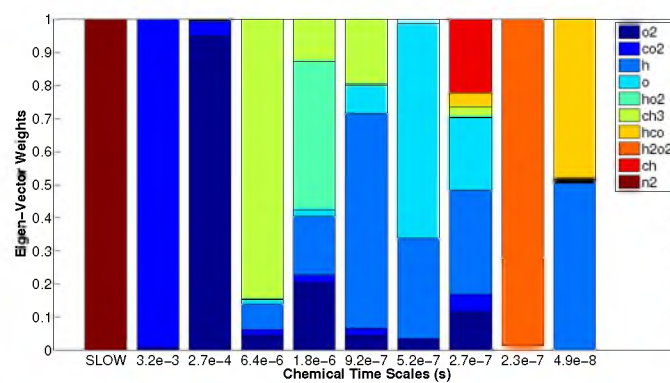


Figure 2.14: Stacked bar chart showing normalized chemical species weights from the eigenvector matrix corresponding to chemical time-scales from the inverse of the eigenvalues.

correspond to radical species where characteristic reaction times are expected to be very small (H , HCO , H_2O_2 , O , HO_2 , CH).

A simple comparison between the newly outlined approach for calculating τ_c and the global chemistry approach [40] can be made using the one-step Westbrook and Dryer mechanism [55], assuming the oxidation of CH_4 is the slower predominant chemical process in the system. The one-step Westbrook and Dryer reaction is given with units of $kcal$, mol , K , m^3 , and s as:



The units for the reaction rate constant are $1/s$, leaving a simple expression which gives τ_c . Figure 2.15 shows the results for the one-step global reaction time-scale (blue ‘+’ markers) in comparison to the PV approach (black ‘x’ markers) both calculated from the HM1 data-set. A similar value for the smallest scale time-scale near the stoichiometric mixture fraction ($\xi_{st} = 0.05$) is observed, with large differences moving away from ξ_{st} in either direction. The simulation results for HM1 were recalculated using EDC with the simple one-step reactions for CH_4 and H_2 combustion. The chemical time-scale analysis now shows a large reduction in τ_c (red ‘diamond’ markers) when using the simplified chemical reaction scheme in attempt to model the diluted combustion. Figure 2.15 points out that more detailed chemistry is needed to correctly account for the turbulence-chemistry interaction in the HM1 system. Using EDC in combination with one-step chemistry leads to overestimation of chemical time-scales, and over-prediction of temperature and reaction rates.

In order to verify the proposed methodology, the DNS data-set (described in the Test Cases Section) is also analyzed using the new chemical time-scale approach. First τ_c is calculated using the full Jacobian matrix containing the information from all of the variables, including that which may produce slow dormant reaction times. The flow time-scale τ_f is calculated as either the Kolmogorov time-scale $\tau_\eta = (\frac{\nu}{\varepsilon})^{1/2}$, or the integral time-scale $\tau_I = \frac{k}{\varepsilon}$ where ε , the energy dissipation rate, is calculated from the velocity gradients. Figure 2.16 shows the results using the full Jacobian matrix.

Similar to the analysis shown in Figure 2.3, the dormant reaction times hide the actual governing time-scales of the system. In order to perform the proposed approach, Equation 3.18 was used again to determine the number of principal variables required for sufficient description of the data-set. Figure 2.17 shows the normalized trace of the variance which is lost based on a given value of q , where $q \geq 2$ will yield a 1% or less loss of variance. Figure 2.18 shows τ_c and Da_η calculated for the given DNS data-set. The analysis showed a clear definition of the dominant reaction time-scale while using one to eight principal variables.

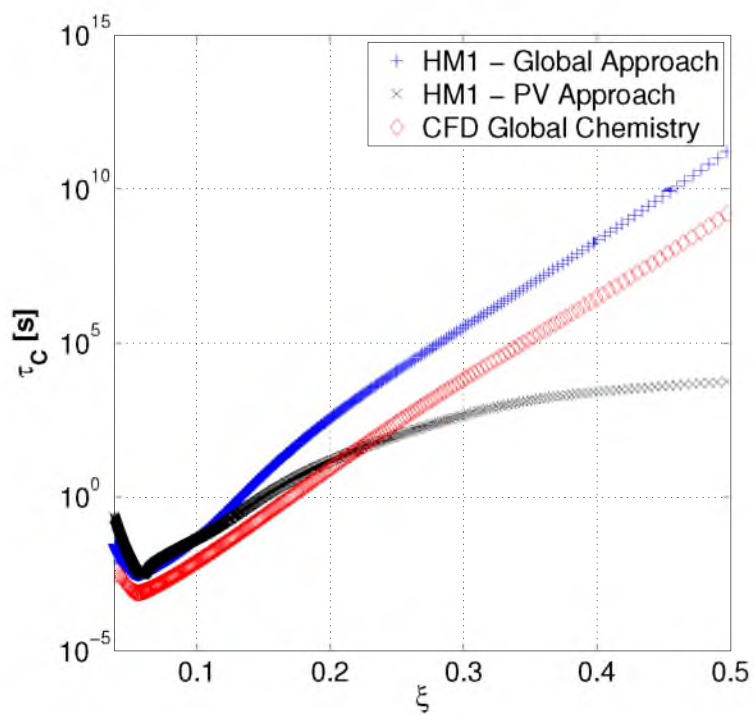
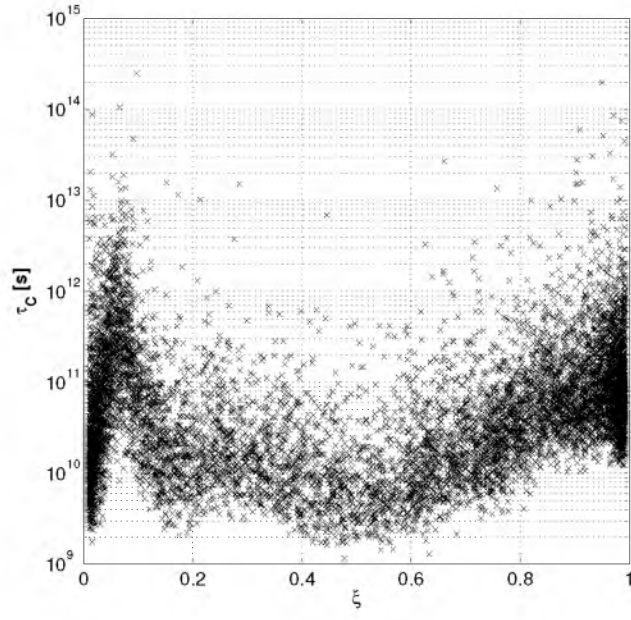
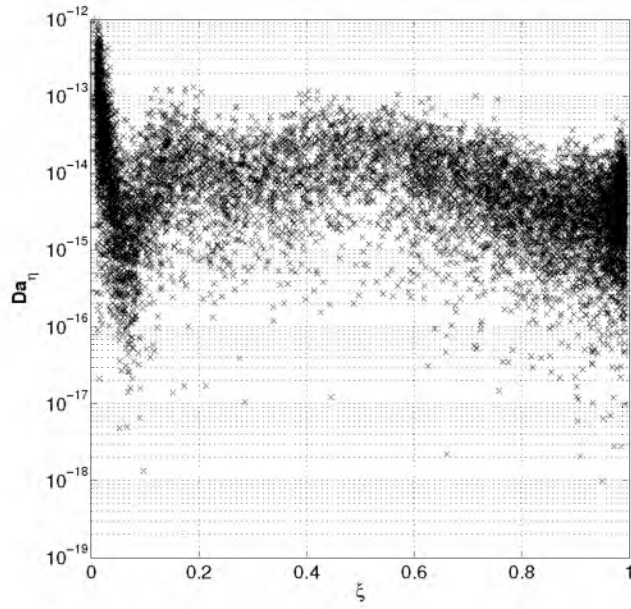


Figure 2.15: Comparison of chemical time-scales τ_c for the HM1 data-set calculated using the principal variable approach (black 'x' markers) or with global chemistry (blue '+' markers) and τ_c from the the calculation of HM1 using global chemistry with the EDC model (red 'diamond' markers).



(a)



(b)

Figure 2.16: Chemical time-scale τ_C (a), and Damköhler Da_η (b) values as a function of mixture fraction (ξ) using the full Jacobian matrix.

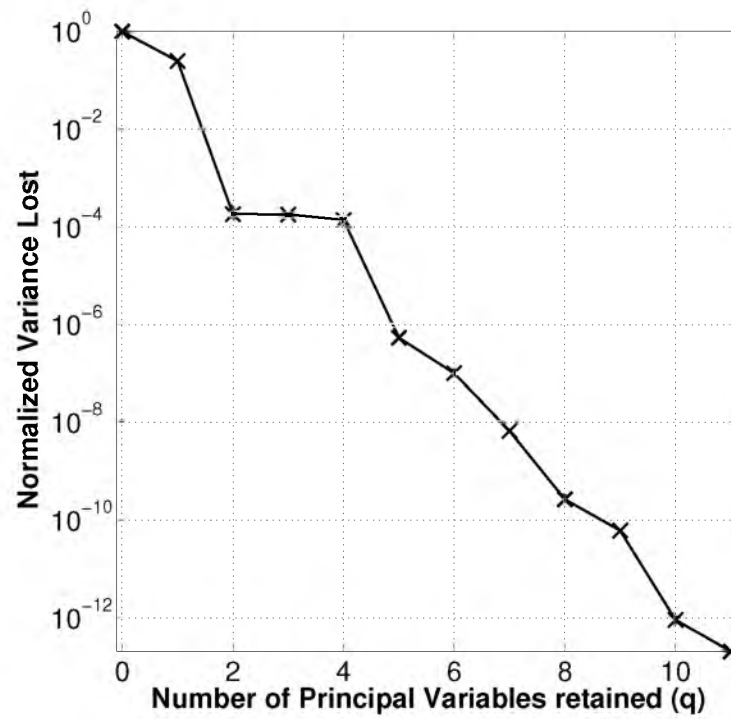
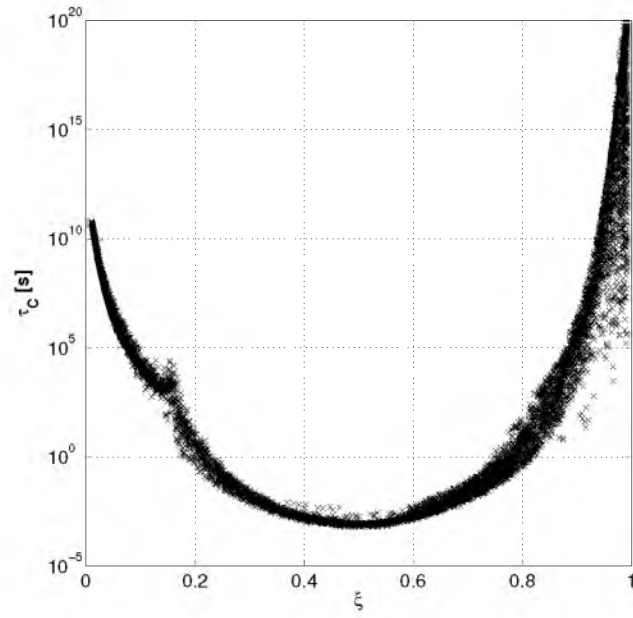
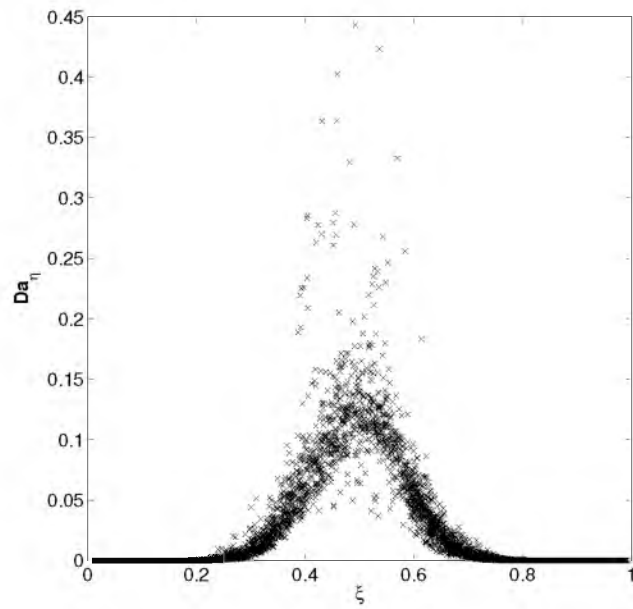


Figure 2.17: Trace plot for DNS case. Y-axis gives a normalized variance which is lost based on the selection of q , the x-axis.



(a)



(b)

Figure 2.18: Chemical time-scale τ_C (a), and Damköhler Da_η (b) values as a function of mixture fraction (ξ), using two principal variables.

Upon addition of the ninth principal variable, which was H_2O in this case, dormant reaction times are observed. Figure 2.19 shows a range of Damköhler numbers comparing Da_I and Da_η . Here we observe a large integral Damköhler number, thus a flamelet-like system is expected (and observed), yet a Kolmogorov Damköhler number less than unity leaving the nonpremixed reaction zones with enhanced mixing as the time-scales of the smallest turbulent structures are smaller than the limiting reaction times, allowing mixing within the reaction zones. The decrease in lost variance (shown in Figures 2.6 and 2.17) is much faster for the CO/H_2 data-set. In particular, the system has a limited degree of extinction and re-ignition, leading to a simpler structure and minor influence of finite rate chemistry effects, requiring a smaller number of PV (when compared with the CH_4/H_2 data-set). In addition, larger fuels will inherently require more PVs, as seen when comparing the two data-sets.

2.6 Conclusions

A method has been described for the calculation of chemical time-scales for turbulent combustion data. The method identifies the limiting chemical time-scales by identifying the principal variables of the system, and using them in the evaluation of the chemical source term Jacobian. With a relevant definition of a mixing time-scale, the Damköhler number may be calculated. The Damköhler number is useful in helping to identify combustion regimes and appropriate modelling strategies. The presented method can be used in turbulent flame studies in the following manner: First, a high fidelity numerical data-set representative or relevant to a system of interest is generated and a PV analysis is performed to extract the leading time-scales of the systems, using the proposed methodology based on the eigenvalue decomposition of the down-scaled chemical Jacobian. The extracted chemical time-scales are then compared to the mixing time-scales for the estimation of the Damköhler number. A large Damköhler number corresponds to combustion regimes dominated by mixing, indicating that flamelet-like modelling strategies are appropriate. On the other hand, in case of slower chemistry or enhanced mixing, lower Damköhler numbers are found and more advanced modelling strategies are required, in order to adequately describe turbulence-chemistry interactions and finite-rate chemistry effects.

2.7 Acknowledgments

We thank Dr. Alberto Cuoci for his assistance in the development of the JACOBEN software. We are grateful to the reviewers who provided valuable feedback. In addition we thank our sponsor for which part of the present research was funded: The National Nuclear

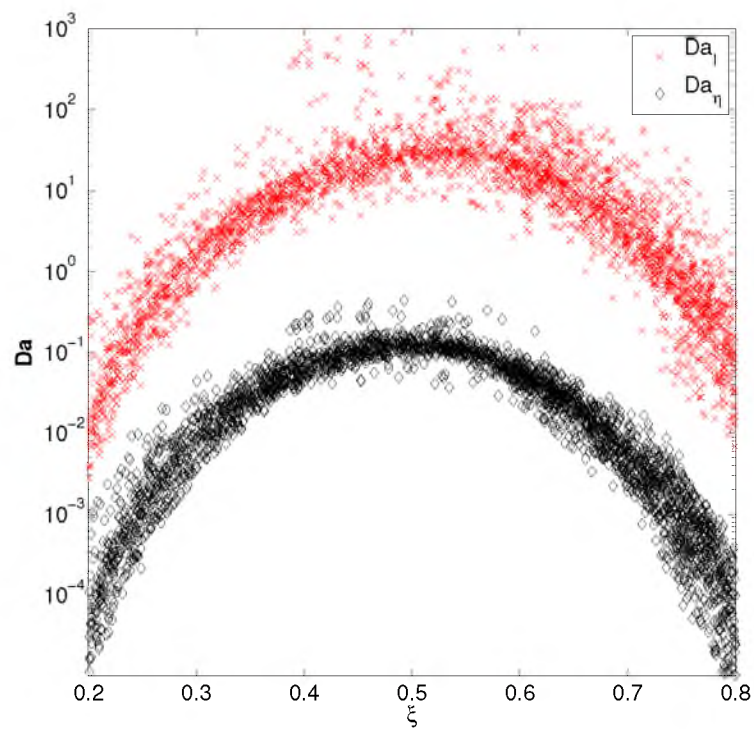


Figure 2.19: Damköhler numbers calculated using the integral time-scale and the Kolmogorov time-scale using two principal variables

Security Administration under the Accelerating Development of Retrofittable CO₂ Capture Technologies through Predictivity program through DOE Cooperative Agreement DE NA 00 00 740.

CHAPTER 3

REDUCED-ORDER PCA MODELS FOR CHEMICAL REACTING FLOWS

Benjamin J. Isaac, Axel Coussement, Olivier Gicquel, Philip J. Smith, and Alessandro Parente

Revised version submitted to Combustion and Flame, February 2013

3.1 Abstract

One of the most challenging aspects of turbulent combustion research is the development of reduced-order combustion models which can accurately reproduce the physics of the real system. The identification and utilization of the low-dimensional manifolds in these system is paramount to understand and develop robust models which can account for turbulence-chemistry interactions. Recently, principal components analysis (PCA) has been given notable attention in its analysis of reacting systems, and its potential in reducing the number of dimensions with minimum reconstruction error. The present work provides a methodology which has the ability of exploiting the information obtained from PCA. Two formulations of the approach are shown: Manifold Generated from PCA (MG-PCA), based on a global analysis, and Manifold Generated from Local PCA (MG-L-PCA), based on performing the PCA analysis locally. The models are created using the co-variance matrix of an empirical data-set which is representative of the system of interest. The reduced models are then used as a predictive tool for the reacting system of interest by transporting only a subset of the original state-space variables on the computational grid and using the PCA basis to reconstruct the nontransported variables. The present study first looks into the optimal selection of the subset of transported variables and analyzes the effect of this selection on the approximation of the state-space and chemical species source terms. Then, a demonstration of various *a posteriori* cases is presented.

3.2 Introduction

It is well-established that the ability to model industrial combustion systems is dependent on the ability to represent the reaction system with a reduced number of parameters. Literature enumerates numerous approaches to reduce the computational cost associated with turbulent combustion problems. Several methods are based on the parameterization of the state-space with a reduced number of optimal variables. This leads to fewer transport equations, and provides a reduction in computation time. Several examples include: Steady Laminar Flamelet Method (SLFM) [67, 66], Flamelet-Generated Manifold (FGM) [56, 88] and Flamelet-Prolongation of ILDM model (FPI) [27, 21].

Recent work has been pushing for the development of a new class of models which are entirely based on empirical data-sets. The concept is to use principal components analysis (PCA) on empirical data-sets to identify a low-dimensional representation of the reacting system. Previous work by Maas and Thévenin [42] applied PCA to premixed DNS cases, to identify correlations between species concentrations. In the work by Parente et al. [61, 60], PCA was used to identify the best linear representation of the underlying manifold contained in these highly coupled reacting systems. Biglari and Sutherland [5] and Pope [72] extended such a concept using the PCA basis in conjunction with nonlinear regression, maximizing the size reduction for a given accuracy. Mirgolbabaei and Echehki [50] extended such an analysis to the application of artificial neural networks, showing also the effect of minor species on the accuracy in the reconstruction. Mirgolbabaei and Echehki [49] investigated the potential of kernel PCA, showing the high compression potential derived by transforming the initial problem into a nonlinear featured space where linear PCA is carried out. These works show the capability of PCA to recover a highly accurate reconstruction of the state-space variables with a significant dimension reduction. This indicates that a lower dimensional manifold exists in turbulent reacting systems, and that PCA is well-suited for identifying the manifold.

Various approaches have been developed in order to use the manifold identified by PCA. The PC-score approach was first described by Sutherland and Parente [85] as a model which transports the principal components (PCs) directly. Several other groups have proposed transporting a subset of state-space variables and reconstructing the nontransported variables from the PC basis [16, 91]. In particular, the MG-PCA model by Coussement et al. [16] was the first PCA model for which *a posteriori* validation was provided, by computing a hydrogen flame-vortex interaction using a DNS solver. Finally, in the work by Najafi-Yazdi et al. [52], PCA was used to identify optimal progress variables in the context of the

flamelet-generated manifold approach.

The present paper focuses on the use of PCA for combustion models. Primarily, the Manifold Generated from PCA (MG-PCA) model [16] is examined. In this model, transport equations are solved for a subset of the originally transported variables which contain most of the variance of a reacting system. The remaining variables are reconstructed from the PCA basis which has been calculated *a priori*. The current work aims at extending the analysis of the MG-PCA method by proposing an enhancement to the model to increase its accuracy, investigating various *a priori* aspects of the MG-PCA models, and by showing two *a posteriori* demonstrations of the model with a more challenging chemistry than in [16]. The *a priori* investigation shows the improvement in accuracy provided by the new model formulation, the effect of the various transported variables selection methods and preprocessing techniques on size reduction, as well as on reconstruction of state-space variables and source terms. Then two MG-PCA syngas (CO/H_2 mixture) calculations are shown, including a constant pressure auto-ignition case, and a turbulent syngas premixed flame.

3.3 Principal component analysis

For a data-set $\mathbf{X} (n \times Q)$, containing n samples of Q original variables, PCA provides an approximation of the original data-set using only q ($q < Q$) linear correlations between the Q variables [85, 61]. PCA starts with the computation of the sample co-variance matrix \mathbf{S} :

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} \quad (3.1)$$

where the superscript T indicates the transpose matrix. Using the spectral decomposition, \mathbf{S} is then decomposed to:

$$\mathbf{S} = \mathbf{A} \mathbf{L} \mathbf{A}^T \quad (3.2)$$

where $\mathbf{A} (Q \times Q)$ and $\mathbf{L} (Q \times Q)$ are respectively the Q eigenvectors of \mathbf{S} , called principal components (PCs), and the eigenvalues of \mathbf{S} , in decreasing order. The principal component scores, $\mathbf{Z} (n \times Q)$, are then computed using the eigenvector matrix as:

$$\mathbf{Z} = \mathbf{X} \mathbf{A}. \quad (3.3)$$

One of the main advantages of PCA is that the original set of data (\mathbf{X}) can be uniquely recovered using the PCs and their associated scores:

$$\mathbf{X} = \mathbf{Z} \mathbf{A}^{-1} \quad (3.4)$$

where it should be noted that $\mathbf{A}^{-1} = \mathbf{A}^T$. However, the main objective of PCA is dimension reduction. Indeed, if one only uses the first q PCs ($q < Q$), an approximation of \mathbf{X} based on the first q eigenvectors (\mathbf{X}_q) is obtained:

$$\mathbf{X} \approx \mathbf{X}_q = \mathbf{Z}_q \mathbf{A}_q^T \quad (3.5)$$

where \mathbf{X}_q is the approximation of \mathbf{X} based on the first q eigenvectors of \mathbf{S} , and \mathbf{Z}_q is the $n \times q$ matrix of the principal component scores.

Finally, it should be stressed that throughout this paper, the data are preprocessed prior to performing PCA. In particular, each variable of the original data-set \mathbf{X} is centered and scaled in order to increase the accuracy of the method [61, 63]. Applying centering and scaling on the data-set reads:

$$\mathbf{X}^s = (\mathbf{X} - \overline{\mathbf{X}}) \mathbf{D}^{-1} \quad (3.6)$$

where $\overline{\mathbf{X}}$ is a $n \times Q$ matrix containing the mean of each variable and \mathbf{D} is a $n \times Q$ matrix containing the standard deviation of each variable (see [63] for details).

Using the previous analysis, two general classes of PCA-based combustion models have been identified:

- First, one can directly transport the principal components, as proposed in the work by Sutherland and Parente [85]. The thermo-chemical state-space is then recovered using Equation 3.5. While the approach is straightforward, it suffers from a major drawback related to the PC source terms. In particular, the error associated to the PCA reconstruction strongly affects the calculation of the source terms, whose accuracy degrades very quickly when reducing the number of parameters defining the manifold. This is due to the fact that PCA evenly distributes the reconstruction error on the state variables, without taking into account the absolute size of the variables. As a consequence, radical species present in very small amounts are affected by reconstruction errors of the same order of magnitude as the major variables, leading to an uncontrolled propagation of error when the source terms are calculated from the approximated state-space [5]. Therefore, nonlinear regression techniques are being used to parameterize the full thermo-chemical state.
- Second, one can transport a subset of the original variables and recover the remaining variables using the information from PCA (MG-PCA). The MG-PCA approach was developed [16] to better control the propagation of the reconstruction error. Such an approach is based on the resolution of classic transport equations for the system

principal variables. Indeed, Equation 3.5 indicates that \mathbf{X}_q can be obtained from \mathbf{Z}_q . Moreover, those scores can be approximated from a subset of the original variables $\mathbf{X}(q)$ of size $n \times q$ composed of only q variables:

$$\tilde{\mathbf{Z}}_q = \mathbf{X}(q) (\mathbf{A}(q)_q^T)^{-1} \quad (3.7)$$

where $\mathbf{A}(q)_q$ is a $(q \times q)$ matrix containing only the coefficients related to the q retained variables. Combining Equations 3.5 and 3.7, one finds:

$$\mathbf{X}_q = \mathbf{X}(q) (\mathbf{A}(q)_q^T)^{-1} \mathbf{A}_q^T \quad (3.8)$$

or

$$\mathbf{X}_q = \mathbf{X}(q) \mathbf{B} \quad (3.9)$$

defining the matrix \mathbf{B} ($q \times Q$) as:

$$\mathbf{B} = (\mathbf{A}(q)_q^T)^{-1} \mathbf{A}_q^T. \quad (3.10)$$

Therefore, by transporting q variables (which can be temperature or species mass fractions), it is possible to recover the $(Q - q)$ remaining variables by retaining the appropriate $(Q - q)$ columns of the \mathbf{B} matrix in Equation 3.9, corresponding to the nontransported $(Q - q)$ state variables:

$$\mathbf{X}_q (Q - q) = \mathbf{X}(q) \mathbf{B} (Q - q). \quad (3.11)$$

By comparison with the score approach, this method requires the *a priori* selection of q transported variables. This method allows one to better control the propagation of the error linked to the model, which is the major advantage of MG-PCA.

MG-PCA allows the use of PCA locally [35]. Parente et al. [61] first applied the local PCA formulation to turbulent combustion data, identifying the limitations of global PCA for the analysis of highly nonlinear systems as the ones observed in combustion. In fact, PCA tries to approximate the nonlinear chemical manifold by superimposing several linear effects, resulting in a manifold size higher than the actual problem dimensionality. To avoid such a problem, the local PCA approach was proposed to optimally partition the data into clusters, based on an iterative algorithm which minimized the reconstruction error of the state-space. However, the implementation of local PCA in terms of a combustion model does not appear straightforward for two main reasons: first, the approach is based on the resolution of transport equations for the scores, implying a modification of the PC definition

with the cluster, and, second, the conditioning variable is not known *a priori* and it is not guaranteed that it could be somehow related to any state variable¹. On the other hand, the use of local PCA appears well-suited in the MG-PCA context, as indicated in [16]. The main steps of the approach, briefly indicated as MG-L-PCA, are:

- The principal variables are extracted from the full data-set, to define the transport equations which need to be resolved in all identified clusters.
- Then, the matrices \mathbf{A}_q and \mathbf{B} are computed in each cluster, allowing an optimal local reconstruction of the nontransported variables using Equation 3.5.

Differently from [61], the conditioning variables are chosen *a priori* to build continuous clusters on the basis of a progress variable displaying a monotonic increase throughout the flame. For premixed cases, as the ones described in the present paper, temperature represents an optimal choice² [16].

The MG-PCA algorithm can be divided into two parts:

- First, the data-set \mathbf{X} is generated using a "canonical reactor" with the same chemical composition of the system to be simulated. Obviously, the data-sets should be simple to compute, in order to generate combustion models of tailored-accuracy with affordable computational resources. In the present work, a one-dimensional premixed flame is used; however, for nonpremixed systems, steady laminar flamelets [66] with varying strain-rate could be used.
- A principal component analysis is then performed to identify the manifold. The \mathbf{B} matrix is computed and the subset of q retained variables is identified. Note that if the local formulation is used, one has to identify the clusters and compute their corresponding \mathbf{B} matrices.

Then the database is used in the flow solver:

- Transport equations are solved for the q nonconserved scalars.
- At the end of each temporal (or pseudo-temporal) iteration, the missing $(Q - q)$ variables are reconstructed using the \mathbf{B} matrix (Equation 3.11) at every grid point.

¹In the context of nonpremixed flames, it was shown that the conditioning variable corresponded quite well to mixture fraction [61]

²For nonpremixed flames, mixture fraction would represent an optimal choice, as indicated in [61].

- All the species are then available for the next temporal (or pseudo temporal) iteration. The diffusion and source terms appearing in the conservation equations of the q retained variables are computed using a CHEMKIN-like [36] formalism for all the species (retained and recovered).

The \mathbf{B} matrix and the coefficients used to center and scale the data are constants³. Therefore, this algorithm requires only one matrix-vector multiplication. The computational cost to recover nontransported species using the MG-PCA technique is therefore very low.

3.4 Challenges of the MG-PCA model

Before applying the MG-PCA model to actual computations, several issues must be carefully addressed:

- A common issue in PCA-based models is the need for an empirical data-set which represents the system of interest. The data-set also needs to be easy to compute (e.g. 1D flame solutions).
- The accuracy of the model will obviously rely on the accuracy of the reconstruction of the missing variables. Since the method relies on the matrix \mathbf{B} for the reconstruction of the nontransported variables, its computation must be as accurate as possible. The use of Equation 3.10 does not provide satisfactory results, as it will be shown below, and an alternative optimal estimation of \mathbf{B} must be provided for the success of the method.
- The selection of the transported variables is crucial, requiring that the most informative variables in the data-sets be selected. Moreover, the number of transported variables, q , must be appropriately selected to ensure that the source terms are accurately reconstructed for the reduced set of scalar transport equations
- The scaling and centering coefficients (see Equation 3.6) have a great impact on the accuracy of the method and they must be chosen with care.

3.4.1 Reference data-set

In order to address the issues presented above, an *a priori* demonstration is now provided to cover the aforementioned aspects using a DNS flame-turbulence data-set. The

³Note that if local-PCA is used, the coefficient for centering and scaling along with the \mathbf{B} matrix are dependent on the cluster which must first be identified [16].

compressible flow solver YWC, developed at the EM2C Laboratory by Coussement et al. [15, 17], is used to generate this data-set. The 2D flame turbulence field is initialized using a 1D flame extended along the y -axis and by super-imposing a turbulence field. The turbulence field is initialized using the Passot-Poquet spectrum [64] with $Re_t = 1423$ giving a Kolmogorov length scale of $l_k = 8.2 \cdot 10^{-6}$. The initial turbulence field is shown in Figure 3.1. The computational domain extends $8 \cdot 10^{-3}$ m in both x and y direction, and mesh spacing is $1.25 \cdot 10^{-6}$ m. The fuel considered is syngas (CO/H_2 mixture, 50/50 molar basis), the oxidizer is air, and the equivalence ratio is $\phi = 0.88$. Boundary conditions consist in an inlet at $x = x_{min}$ which imposes the turbulent field generated with a convection velocity of 4 m/s along the x -axis. At $x = x_{max}$, an outlet boundary is imposed with $p = 101,325$ Pa. Remaining boundary conditions ensure periodicity along the y -axis. The chemical scheme is the one from Davis et al. [20] including 12 species ($N_s = 12$): N_2 , O_2 , H_2 , H_2O , H_2O_2 , CO , CO_2 , O , H , OH , HO_2 , and HCO . Thermo-chemical and transport properties are computed using a CHEMKIN-like formalism [36]. The DNS field used to perform the analysis below is taken at $t = 7.118 \cdot 10^{-4}$ s. Mass fraction of H_2O (Y_{H_2O}) and HCO (Y_{HCO}) are given in Figure 3.2.

It is important to note that differential diffusion is used. With $N_s = 12$ and only a constraint on mass conservation (here energy and elemental mass fractions are not constant), there are N_s degrees of freedom, because pressure is also constant ($p = p_{atm}$). Therefore, a true reduction of the state-space dimensionality is achieved if the proposed methods allow one to transport less than $q = 12$ variables (if energy is included in the PCA analysis), while maintaining accuracy.

3.4.2 Computation of the \mathbf{B} matrix

The computation of the \mathbf{B} matrix using Equation 3.10 is not optimal. To illustrate this, we revisit the definition of the scores computed using Equation 3.7:

$$\tilde{\mathbf{Z}}_q = \mathbf{X}(q) (\mathbf{A}(q)_q^T)^{-1}.$$

Depending on the number of retained principal variables contributing to the definition of $\mathbf{A}(q)_q$, the approximated scores ($\tilde{\mathbf{Z}}_q$) can result in a weak representation of the state-space. This constrains the achievable reduction when Equation 3.8 is employed. However, being that the MG-PCA model is built *a priori*, this limitation can be overcome by considering the real q principal component scores, obtained using the first q components of Equation 3.3:

$$\mathbf{Z}_q = \mathbf{X} \mathbf{A}_q. \quad (3.12)$$

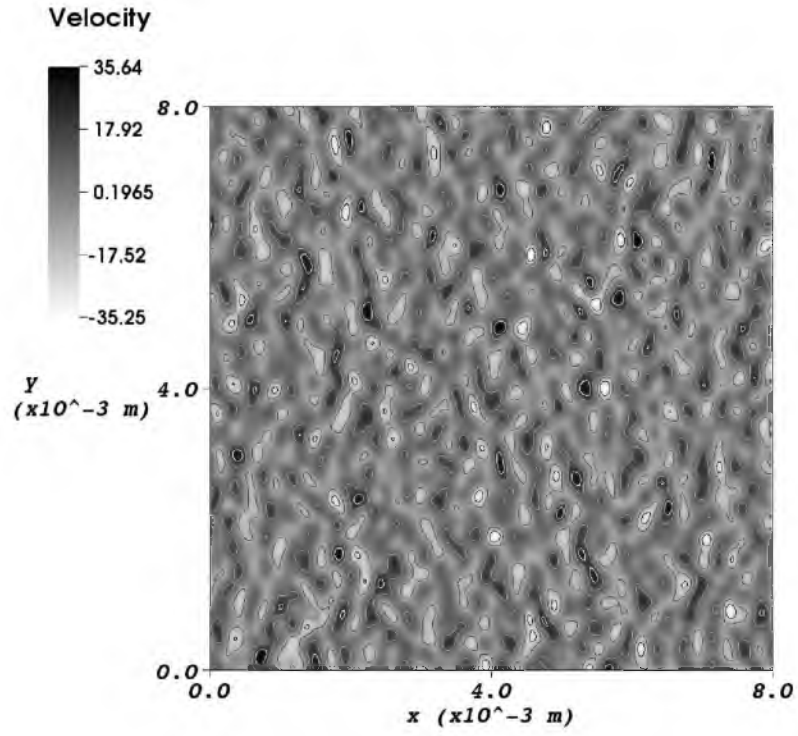


Figure 3.1: Initial turbulence field for the 2D DNS, x-component velocity

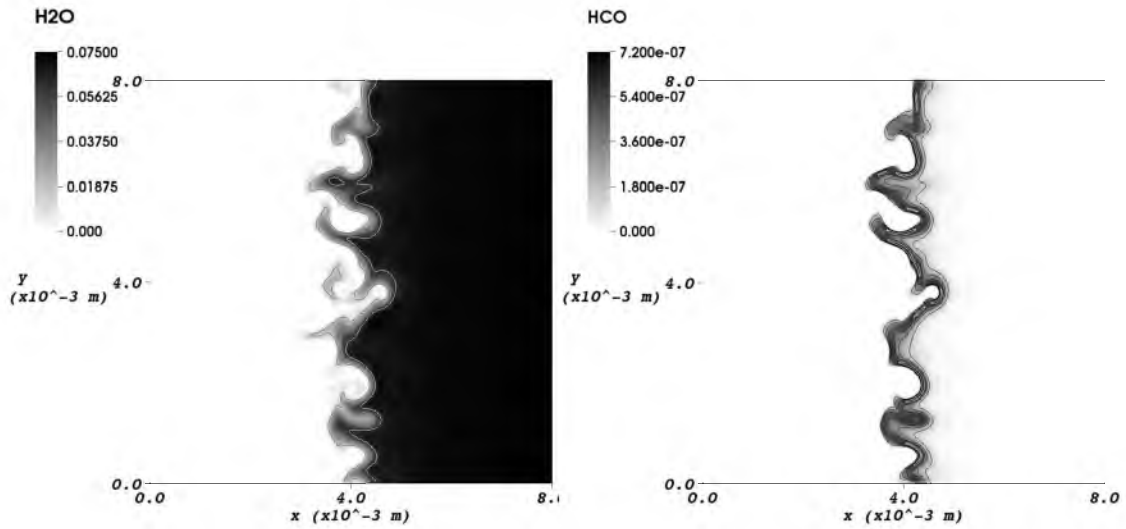


Figure 3.2: Y_{H_2O} , Y_{HCO} field for the 2D DNS, $t = 7.11799 \times 10^{-4}$ s.

Here all of the state-space variables in \mathbf{X} as well as all of the weights in the first q eigenvectors are used instead of a subset. Thus, by using \mathbf{Z}_q instead of $\tilde{\mathbf{Z}}_q$, there should be a general decrease in reconstruction error in the nontransported species. In order to reflect this change, the calculation of \mathbf{B} is performed using:

$$\mathbf{B} = \mathbf{X}(\mathbf{q})^+ \mathbf{Z}_q \mathbf{A}_q^T \quad (3.13)$$

which represents the solution in the least squares sense to the system given by Equation 3.12. The matrix $\mathbf{X}(\mathbf{q})^+$ is the pseudo inverse of $\mathbf{X}(\mathbf{q})$ of size $q \times n$, which is given by

$$\mathbf{X}(\mathbf{q})^+ = \left(\mathbf{X}(\mathbf{q})^T \mathbf{X}(\mathbf{q}) \right)^{-1} \mathbf{X}(\mathbf{q})^T. \quad (3.14)$$

The two methods for the calculation of \mathbf{B} are compared in Figure 3.3 with reference to the 2D DNS data-set. The approximation error is reported using a normalized root mean square distance identical to that used by Pope [72], defined as:

$$nrms_j = \frac{1}{std(\phi_{ij})} \left(\left(\frac{1}{N} \sum_{i=1}^n (\phi_{ij} - \tilde{\phi}_{ij})^2 \right)^{1/2} \right). \quad (3.15)$$

The *nrms* error statistic for the nontransported state variables are calculated, and the largest *nrms* value is plotted. The figure illustrates the improvement in accuracy.

3.4.3 Selection of the transported variables

In the current approach, the method used to select the q transported variables which are transported relies on the principal variables concept [30]. Principal components (PC) are linear combinations of all the variables defining the data-set. However, these variables are not necessarily equally important to the formation of the PCs. Some of the variables may be critical whereas others may be redundant. Motivated by this fact, one can try linking the PC back to a subset of the original variables, which satisfy one or more optimal properties of PCA, such as the maximization of the variance of the original data \mathbf{X} . A number of methods exist for selecting a subset of q original variables. The following methods are considered here:

- *B2 backward and B4 forward methods* [32]. Variables are determined by analyzing the principal component weights (\mathbf{A}_q). The *B2 backward method* removes variables associated to the PC with the smallest eigenvalues. In contrast, the *B4 forward method* identifies variables which are associated with the PC with the largest eigenvalues.

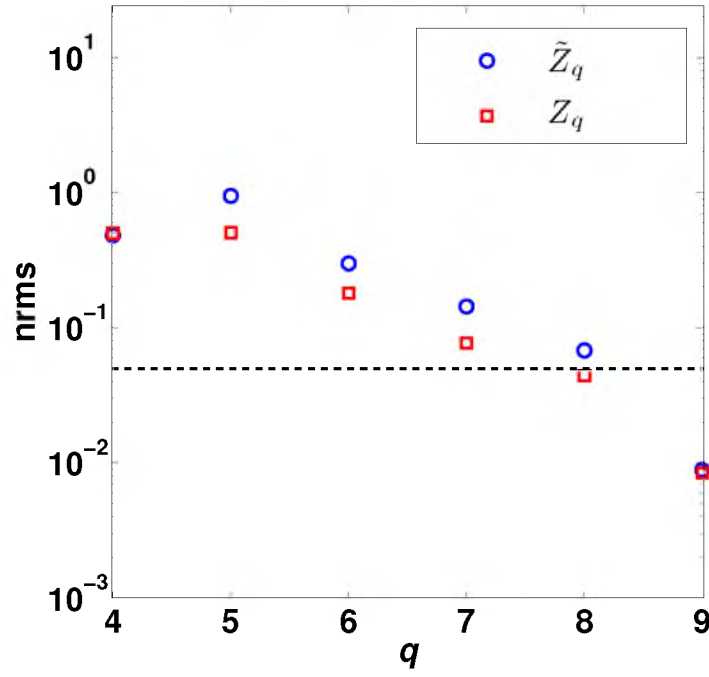


Figure 3.3: Maximum *nrms* distance (y-axis) for the reconstruction of the state-space variables as a function of the number of variables (q) (x-axis), while using the $\tilde{\mathbf{Z}}_q$ or \mathbf{Z}_q in the construction of the \mathbf{B} matrix. The dashed line shows 5% *nrms*. Scaling: standard deviation. PV selection: B2.

- *M2 backward method* [39]. Variables are determined by comparing \mathbf{Z} (Equation 3.3) with an approximate $\tilde{\mathbf{Z}}$. Here $\tilde{\mathbf{Z}}$ is constructed from a subset of the original variables, and the comparison is made to the original scores using a Procrustes analysis. Upon selection of appropriate variables, the underlying structure of the data is preserved and the PVs are identified as the subset of variables which have been used for the approximation.
- *McCabe criteria* [48]. McCabe identified that the PCs satisfy a certain number of optimality criterion. The criterion are based on partial co-variance matrices calculated by selecting subsets of variables. In the current study, the *MC1* and *MC2* criterion are used, which rely on the determinant (*MC1*) or trace (*MC2*) of the partial co-variance matrices.
- *Principal features* [13]. Variables are identified by analyzing the correlation between variable weights in \mathbf{A}_q . Variables are grouped according to their correlation, and the *k-means* algorithm [84] is used to extract a given number of variables from each subset, thus providing a representation of each of the groups.

Next, the influence of the different PV approaches on the reconstruction of the state variables is assessed and discussed. A useful metric for this task is one that describes the amount of variance lost by the selection of q (the number of transported variables) [30]. One can partition the variables into 2 groups, one for retained variables and the second for the remaining variables. The sample co-variance matrix can then be calculated:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \quad (3.16)$$

and the partial co-variance can then be given as:

$$\mathbf{S}_{22,1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}. \quad (3.17)$$

Taking the trace of the partial co-variance gives a quantitative value for the amount of variance lost (λ) for a given number of principal variables:

$$\lambda = \frac{\text{trace}(\mathbf{S}_{22,1}(q))}{\text{trace}(\mathbf{S})} \quad \text{for } q = 1, 2, \dots, Q. \quad (3.18)$$

For example, if $q = 0$ then $\lambda = 1$, meaning all variance is lost, or when $q = Q$ then $\lambda = 0$, which means all of the original variance in the system is explained.

Figure 3.4 shows the resultant percentage of lost variance calculated from Equation 3.18, for the various PV selection methods. The figure shows that for $q = 6$ only 1% of the

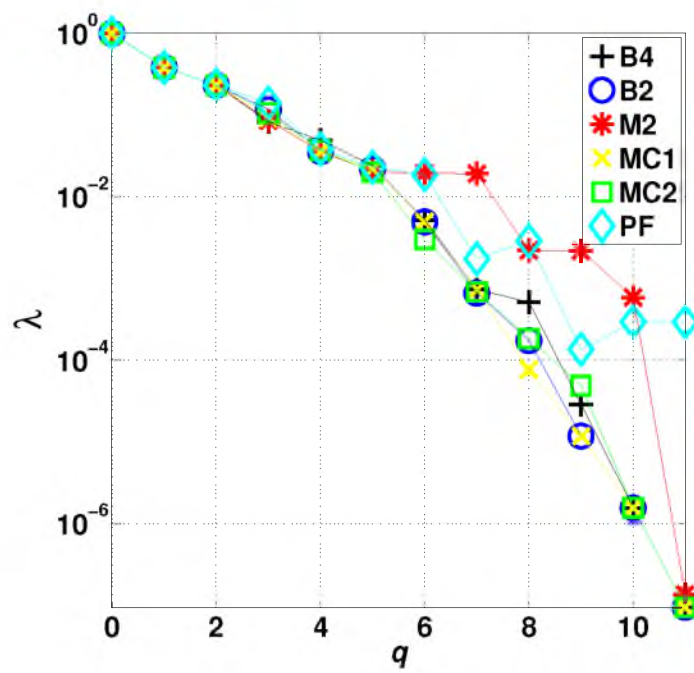


Figure 3.4: Lost variance (y -axis) while adding principal variables (x -axis). The trace is given for the various PV selection methods highlighted in Section 3.4. Scaling: standard deviation

variance is lost while using the B4, B2, MC1, and MC2 methods; the PF, and M2 methods do not achieve such a degree of lost variance until $q = 7$. It is also interesting to note that several of the PV selection methods (M2, and PF) can indeed lose variance upon addition of PVs. The B2, MC1, and MC2, method appear to show a consistent increase in explained variance while adding principal variables. Table 3.1 shows with the acronym *pv* for the principal variables selected by the different approaches, while using auto scaling (see Section 3.4.4), and $q = 7$. It can be observed that different sets of PV are identified, depending on the selection method. However, a common logic seems to hold for all cases: all methods tend to select the majority of the principal variables among the radical species (*HCO*, *HO2*, *OH*, and *H2O2*) which identify ignition or reaction regions, whereas temperature (which is forced to be a principal variable) and one other major variable are in general sufficient for capturing slower changes in the system. The first benchmark among the sets of PV must be carried out with respect to their ability of accurately reproducing the nontransported state-space variables. Table 3.1 lists the *nrms* values for all the state variables reconstructed using MG-PCA and the different PV selection approaches. It can be observed that almost all methods provide a good approximation of the state-space, with an exception for M2 and PF methods, which cannot properly recover some of the radical species (*OH*, *H*, *HO2*). This can be explained by the fact that the methods that perform well select only one of the major species while keeping several radical species, which are usually associated with increased nonlinearity. It is interesting to note that the forward and backward methods (B4 and B2) both select either of the major reaction products *H2O* and *CO2*, whereas the McCabe methods keep either *H2O* or *O2*. The PV and M2 methods performed the worst yielding *nrms* values greater than or equal to 10^{-1} for two or more of the variables. These methods identified two or three major species as PVs, thus decreasing the models ability to represent the minor species.

3.4.4 Centering and scaling

In order to optimally choose the scaling and centering coefficient, several methods will be investigated. Before presenting the methods, it is useful to rewrite Equation 3.6 in a scalar form for the sake of clarity:

$$x_j^s = \frac{x_j - \bar{x}_j}{d_j} \quad \text{for } j = 1, \dots, Q. \quad (3.19)$$

The centering and scaling coefficients, \bar{x}_j and d_j are stored to be used in the prediction step. The following scaling methods are adopted in the present work [63]:

- *auto scaling*, which adopts the standard deviation as the scaling factor, $d_j = s_j$;

Table 3.1: *nrms* distance for the reconstructed state-space with $q = 7$ while testing various PV selection methods. The acronym *pv* indicates if the variable is a principal variable. Scaling method: auto scaling.

		B4	B2	M2	MC1	MC2	PF
<i>T</i>	1	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>
<i>O2</i>	2	$10^{-1.9}$	$10^{-2.1}$	<i>pv</i>	$10^{-1.9}$	<i>pv</i>	<i>pv</i>
<i>H2</i>	3	$10^{-1.3}$	$10^{-1.3}$	<i>pv</i>	$10^{-1.3}$	$10^{-1.4}$	$10^{-1.4}$
<i>H2O</i>	4	<i>pv</i>	$10^{-1.9}$	$10^{-1.8}$	<i>pv</i>	$10^{-1.9}$	<i>pv</i>
<i>H2O2</i>	5	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>
<i>CO</i>	6	$10^{-1.8}$	$10^{-1.9}$	$10^{-2.0}$	$10^{-1.8}$	$10^{-1.8}$	$10^{-2.1}$
<i>CO2</i>	7	$10^{-1.7}$	<i>pv</i>	<i>pv</i>	$10^{-1.7}$	$10^{-1.7}$	$10^{-2.0}$
<i>O</i>	8	$10^{-1.1}$	$10^{-1.1}$	<i>pv</i>	$10^{-1.1}$	$10^{-1.1}$	<i>pv</i>
<i>H</i>	9	<i>pv</i>	<i>pv</i>	$10^{-0.9}$	<i>pv</i>	<i>pv</i>	$10^{-1.0}$
<i>OH</i>	10	<i>pv</i>	<i>pv</i>	$10^{-0.9}$	<i>pv</i>	<i>pv</i>	$10^{-1.0}$
<i>HO2</i>	11	<i>pv</i>	<i>pv</i>	$10^{-0.3}$	<i>pv</i>	<i>pv</i>	<i>pv</i>
<i>HCO</i>	12	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>	<i>pv</i>

- *range scaling*, which adopts the difference between the minimum and maximum variable value as the scaling factor, $d_j = \max(x_j) - \min(x_j)$;
- *pareto scaling* [54], which adopts the square root of the standard deviation as the scaling factor, $d_j = \sqrt{s_j}$;
- *vast* (variable stability) *scaling* [37], which adopts the product between the standard deviation and the coefficient of variation (s_j/\bar{x}_j) as the scaling factor, $d_j = s_j^2/\bar{x}_j$;
- *level scaling*, which adopts the mean value as the scaling factor, $d_j = \frac{1}{N} \sum_{i=1}^N x_j$.

The B2 selection method presented some of the most promising results for the selection of PVs; accordingly, an analysis is now given on the effects of scaling given that the transported variables are selected using the B2 method. Table 3.2 shows the effect of scaling on the reconstruction of the state variables for the DNS case. It is clear from Table 3.2 that the scaling methods have a significant effect on the ability to reconstruct the nontransported variables. The scaling results for the DNS case suggest that auto, range, and pareto scaling provide the most accurate results.

While analyzing the trace using Equation 3.18, Figure 3.5 shows nearly consistent decay in energy upon addition of variables over the various scaling methods, except for pareto scaling. This is due to the very large weight given by such a scaling to temperature [63] with respect to other variables. Thus, the scaled data (under pareto) attribute almost all of the variance in the system to temperature. Temperature is the first PV so the lost variance description is consistent with what is observed in Figure 3.5.

3.4.5 Comparison of the PC-score approach, MG-PCA, and MG-L-PCA

Now a comparison of the three PCA-based modelling techniques is made *a priori* on the 2D DNS field in order to assess their performances. The classic PCA approach presented by Sutherland and Parente [85] has the advantage of not requiring the selection of the systems principal variables, and simply transports the \mathbf{Z}_q of the system (Equation 3.12). This gives an equal distribution of error among all of the state-variables. Table 3.3 shows *nrms* and R^2 error ($R_j^2 = 1 - \sum_{i=1}^n (\phi_{ij} - \tilde{\phi}_{ij})^2 / \sum_{i=1}^n (\phi_{ij} - \bar{\phi}_j)^2$) statistics for the PC-score approach and the MG-PCA approach while attempting to reconstruct the state-space variables with $q = 7$. MG-PCA has difficulty in reconstructing the *O* and *H2* radicals, whereas the PC-score reconstruction is much better overall for the state variables. Table 3.4 lists the R^2 and *nrms* values for the species source terms, which are calculated using the approximate

Table 3.2: *nrms* distance for the reconstructed state-space using $q = 7$ variables, while testing various scaling methods. S = *auto*, R = *range*, P=*pareto*, V=*vast*, L=*level*.

State variables					
	S	R	P	V	L
T	pv	pv	pv	pv	pv
O_2	$10^{-2.1}$	$10^{-2.1}$	$10^{-2.8}$	pv	$10^{-2.0}$
H_2	$10^{-1.3}$	$10^{-1.4}$	pv	$10^{-1.4}$	$10^{-1.3}$
H_2O	$10^{-1.9}$	$10^{-1.8}$	pv	$10^{-1.9}$	$10^{-1.9}$
H_2O_2	pv	pv	$10^{-0.3}$	pv	pv
CO	$10^{-1.9}$	$10^{-2.1}$	$10^{-2.4}$	$10^{-1.9}$	$10^{-1.9}$
CO_2	pv	pv	pv	$10^{-1.8}$	pv
O	$10^{-1.1}$	$10^{-1.1}$	pv	$10^{-1.1}$	pv
H	pv	pv	$10^{-1.0}$	pv	pv
OH	pv	pv	pv	pv	$10^{-1.1}$
HO_2	pv	pv	pv	pv	pv
HCO	pv	pv	$10^{-0.4}$	pv	pv

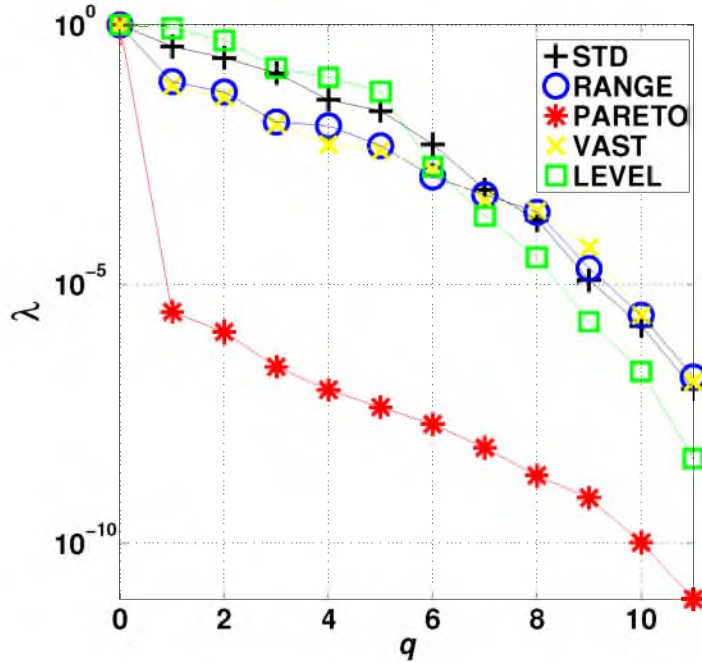


Figure 3.5: Lost variance (y -axis) while adding principal variables (x -axis). The trace is given for the various scaling methods highlighted in Section 3.4. PV selection: B2.

Table 3.3: *nrms* distance and R^2 statistics for the reconstructed state-space with $q = 7$. Scaling: standard deviation. PV selection: B2.

MG-PCA			PC-Score		
Variable	<i>nrms</i>	R^2	Variable	<i>nrms</i>	R^2
<i>T</i>	<i>pv</i>	<i>pv</i>	<i>O2</i>	$10^{-2.0}$	1
<i>O2</i>	$10^{-2.1}$	1	<i>H2</i>	$10^{-1.5}$	0.999
<i>H2</i>	$10^{-1.3}$	0.998	<i>H2O</i>	$10^{-1.7}$	1
<i>H2O</i>	$10^{-1.9}$	1	<i>H2O2</i>	$10^{-2.6}$	1
<i>H2O2</i>	<i>pv</i>	<i>pv</i>	<i>CO</i>	$10^{-1.8}$	1
<i>CO</i>	$10^{-1.9}$	1	<i>CO2</i>	$10^{-1.8}$	1
<i>CO2</i>	<i>pv</i>	<i>pv</i>	<i>O</i>	$10^{-1.7}$	1
<i>O</i>	$10^{-1.1}$	0.994	<i>H</i>	$10^{-2.0}$	1
<i>H</i>	<i>pv</i>	<i>pv</i>	<i>OH</i>	$10^{-1.8}$	1
<i>OH</i>	<i>pv</i>	<i>pv</i>	<i>HO2</i>	$10^{-3.5}$	1
<i>HO2</i>	<i>pv</i>	<i>pv</i>	<i>HCO</i>	$10^{-2.2}$	1
<i>HCO</i>	<i>pv</i>	<i>pv</i>			

Table 3.4: *nrms* distance and R^2 statistics for the reconstructed source terms with $q = 7$. Scaling: standard deviation. PV selection: B2.

MG-PCA			PC-Score		
PV	<i>nrms</i>	R^2	PC	<i>nrms</i>	R^2
<i>T</i>	$10^{-0.4}$	0.87	<i>PC 1</i>	$10^{-0.3}$	0.72
<i>H2O2</i>	$10^{-0.8}$	0.98	<i>PC 2</i>	$10^{0.7}$	0
<i>CO2</i>	$10^{-1.3}$	0.99	<i>PC 3</i>	$10^{0.4}$	0
<i>H</i>	$10^{-0.2}$	0.61	<i>PC 4</i>	$10^{-0.2}$	0.65
<i>OH</i>	$10^{-0.9}$	0.98	<i>PC 5</i>	$10^{1.3}$	0
<i>HO2</i>	$10^{0.3}$	0	<i>PC 6</i>	$10^{-0.3}$	0.80
<i>HCO</i>	$10^{-0.5}$	0.90	<i>PC 7</i>	$10^{0.3}$	0

state-space while retaining 7 PVs (MG-PCA) or PCs (PC-score) of the co-variance matrix, i.e. $q = 7$. The source terms for the PCs ($s_z = \frac{R_i A_z}{\gamma_i}$) are also shown. Even though the PC-score state-space analysis shows an accurate reconstruction of all state-space variables (R^2 of nearly 1 for all of the variables), the error statistics for the source terms show very inaccurate approximations. This condition is complicated by the fact that in the PC-score approach, all the source terms are needed and they should all be computed with great precision. However, this requirement can never be fulfilled as a reconstruction error is always present, without any distinction between the transported and nontransported variables. The PC-score approach has a potential for higher compression, because the variance explained by a linear combination of all of the variables is greater than that provided by a subset of optimal variables. However, without a method to resolve the error propagation in the source terms, the approach can be inaccurate.

3.4.6 MG-L-PCA

The initial analysis of the premixed syngas case indicates that MG-PCA can provide the required precision if $q = 7$. Now the MG-L-PCA method is tested against MG-PCA to investigate the potential of the local PCA formulation. By clustering the data according to temperature, the MG-L-PCA method may now be used to attempt to create a better local basis for the reconstruction of the nontransported variables. Figure 3.6 provides a visualization of the error produced while varying q (x-axis) and c , the number of clusters (y-axis). The gray scale in the figure represents the lowest R^2 statistic for the reconstructed state-space variables. The analysis confirms the earlier findings for MG-PCA, with a minimum R^2 of 0.994 while using $q = 7$ and $c = 1$ (this is in fact the MG-PCA model). By adding clusters, one can better identify the local \mathbf{B} matrix and get better reconstruction of the state-space.

Figure 3.7 also confirms that a higher value of q and c are needed to capture the source term with respect to the state-space, due to the error propagation in the nonlinear source terms. Optimization figures, such as Figures 3.6 and 3.7, are very useful for deciding the parameters of the reduced model to be implemented in a CFD solver, as it provides the minimum number of cluster needed to achieve a desired reduction.

From the above analysis, it can be concluded that the application of MG-L-PCA has the potential of providing accurate results while achieving a significant reduction state-space variables.

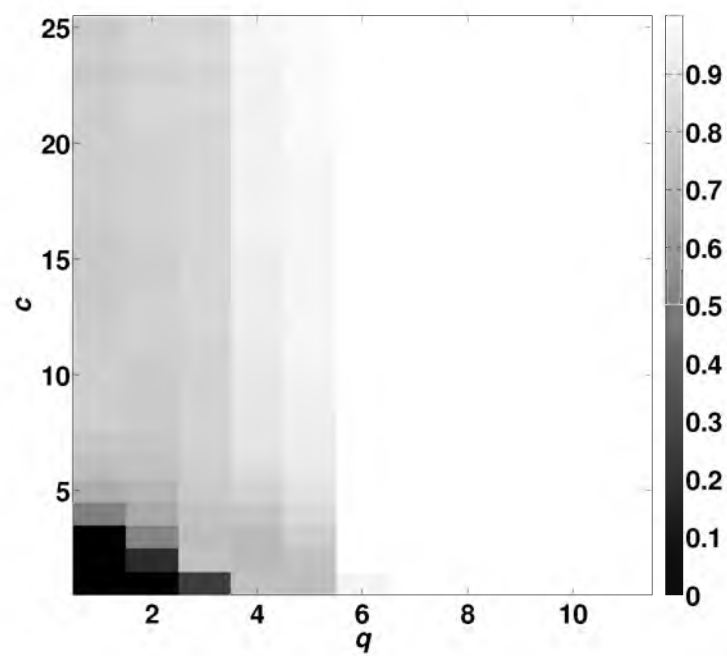


Figure 3.6: Minimum R^2 statistic for the state variables as a function of the number of retained variables (x -axis) and clusters (y -axis). Scaling: pareto. Fuel: Syngas.

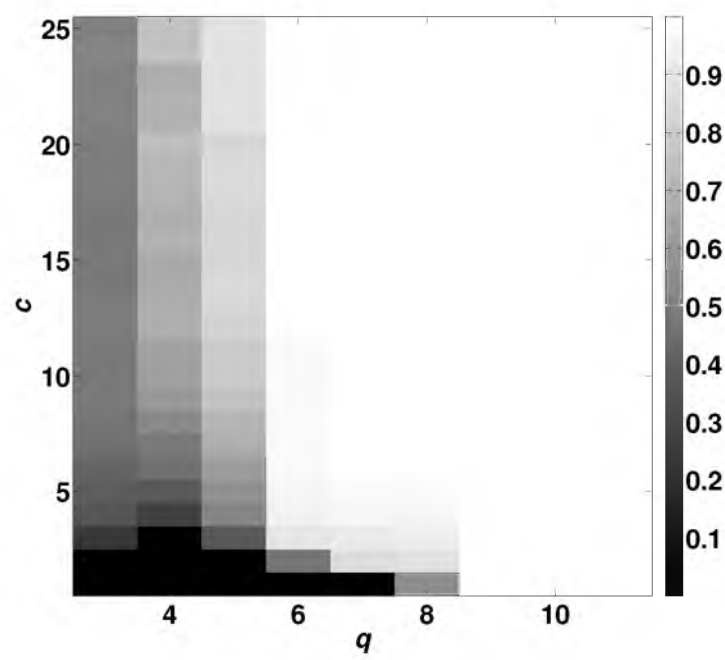


Figure 3.7: Minimum R^2 statistic for the principal variable source terms as a function of the number of retained variables (x -axis) and clusters (y -axis). Scaling: pareto. Fuel: Syngas

3.5 Results

The present section shows actual computations using the MG-PCA and MG-L-PCA methods. The models employ the new definition of \mathbf{B} (Equation 3.13), and a more complicated chemistry, i.e. syngas, than in Coussement et al. [16]. First, an auto-ignition (0D) case is considered with a new technique for cluster identification in the framework of MG-L-PCA. Then, the 2D DNS data-set used in the *a priori* analysis is computed, training the model on a 1D manifold obtained from a laminar premixed flame calculation.

3.5.1 Auto-ignition delay time

The auto-ignition delay time is a rigorous test that demonstrates the ability of a model or a chemical kinetics mechanism to capture complex physical characteristic of the ignition process. The auto-ignition delay time is often tested at various temperatures and pressures, in order to assess the robustness of the model being applied. Accurate prediction of the auto-ignition delay time is particularly important in systems with preheated, and premixed mixtures, such as gas turbines. In the current study, a simplified case is examined with the following assumptions: constant pressure, homogeneous, stagnant premixed fuel, mixture temperature above the auto-ignition temperature of the fuel, and adiabatic conditions. The process is modeled using the differential equations for the chemical species in the system, and the temperature:

$$\frac{dY_i}{dt} = \frac{1}{\rho} W_{s,i} R_i \quad (3.20)$$

$$\frac{dT}{dt} = \frac{1}{\rho c_{P,mix}} \left(- \sum_{i=1}^N h_i W_{s,i} R_i \right) \quad (3.21)$$

where Y_i are the species mass fractions, $W_{s,i}$ is the i^{th} species molecular weight, R_i is the molar source term for the i^{th} species, h_i is the molar enthalpy for species i , and $c_{P,mix}$ is the heat capacity of the mixture. The equations are solved using an in-house implementation of the batch reactor.

In contrast with the DNS case, which exhibits differential diffusion, the degrees of freedom in this case are different. The auto-ignition cases solve 13 transport equations and are constrained by the elemental balances (C, H, O), conservation of mass, and energy, leaving 8 degrees of freedom.

As described in Section 3.4, a data-set is required before calculation in order to determine the number of variables (q) required for a desired accuracy, the variables to transport, and the matrix/matrices \mathbf{B} (MG-PCA/MG-L-PCA). Initially, the full system of equations is calculated at various running conditions in order to generate the data-sets required for the

a priori construction of the model. The number of transported variables q , the transported variables, and the \mathbf{B} matrix/matrices are calculated from each case involving different initial conditions. In the current study, the B2 selection method is used with pareto scaling.

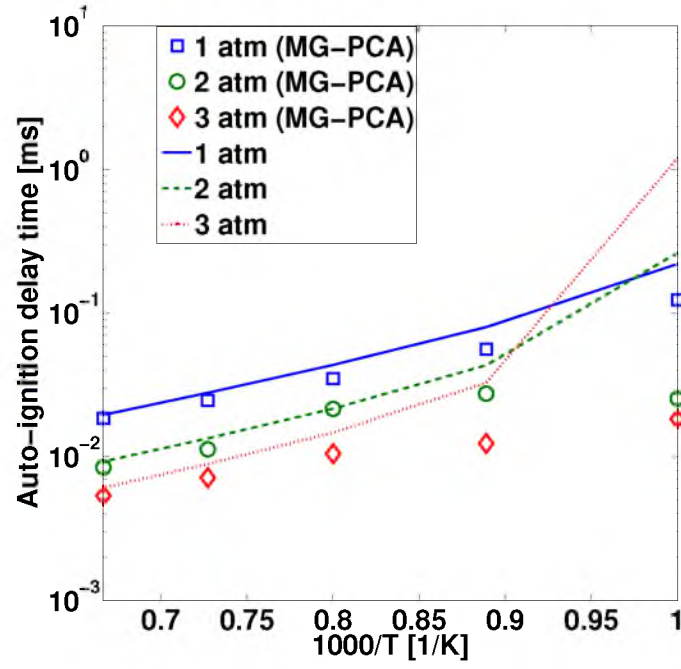
The MG-PCA method shows very good accuracy when using 7 (13% reduction) of the original 12 variables (see Figure 3.8b), which is consistent with the results found in the *a priori* analysis. Minor differences are observed at smaller initial temperatures where the ignition event takes much longer. It is observed that with 6 transported variables that a considerable loss in accuracy occurs over the entire range of pressures and initial temperatures (Figure 3.8a).

While using the MG-L-PCA method, proper clustering of the data is crucial in order to achieve an accurate local reconstruction of the data. As temperature is directly transported in this system, it is an optimal variable for identifying clusters, and at run time providing the local \mathbf{B} matrix which gives the most accurate reconstruction of the local state-space. As one would suspect, problems may arise near cluster boundaries or when a given cluster contains a highly nonlinear peak from one of the transported radical species. Because of this, a clustering algorithm was developed which looks at the *a priori* data and finds local extrema in the radical species profiles, and creates new cluster boundaries at these locations in order to increase the accuracy and provide smoother transition between clusters.

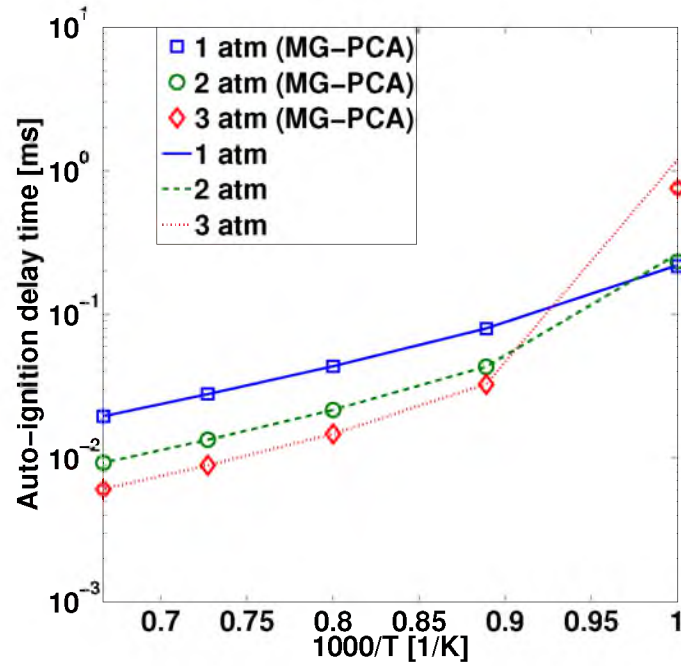
In reference to Figure 3.9b, the results for MG-L-PCA show a much better approximation while transporting as few as 5 variables (38% reduction). However, when moving to 4 variables (50% reduction), reasonable accuracy is observed with moderate initial temperatures, with discrepancies at both higher and lower initial temperatures due to the error from the model (see Figure 3.9a).

3.5.2 Flame-turbulence interaction

The present section reports the result of a DNS calculation of a flame-turbulence interaction case using the MG-PCA method. Numerical setup is exactly the same as in Section 3.4.1 and is not recalled. The simulation is performed using $q = 8$ (33% reduction) in order to reconstruct quasi-exactly the missing variables $R^2 > 0.9999$ for all variables, with the original number of degree of freedom in the system as 12, since differential diffusion is considered and therefore, a reduction of 4 degrees of freedom is achieved. Following the conclusions of the previous sections, pareto scaling is used in combination with the B2 selection method, giving the following variables to be transported inside the solver : T , Y_{H_2} , Y_{H_2O} , Y_{CO_2} , Y_O , Y_{OH} , Y_{HO_2} , Y_{HCO} .

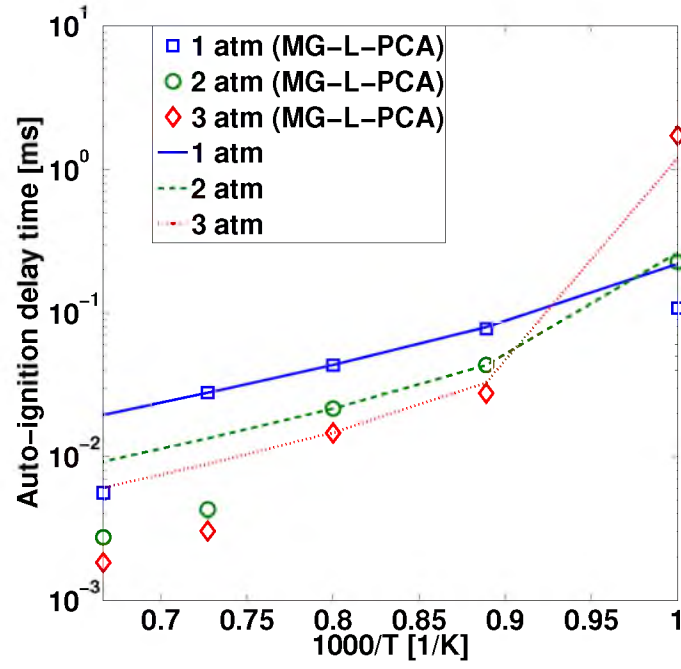


(a)

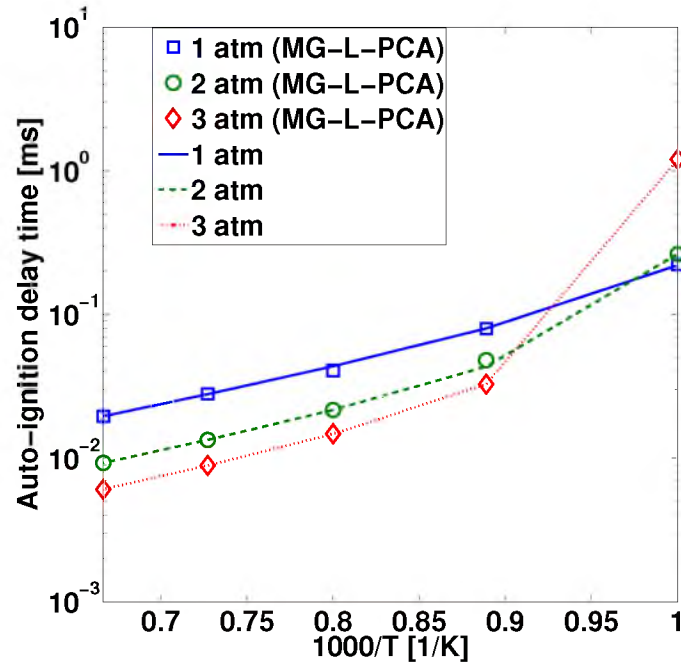


(b)

Figure 3.8: Auto-ignition delay time using MG-PCA as a function of temperature for various system pressures, while $q = 6$ (a) or $q = 7$ (b).



(a)



(b)

Figure 3.9: Auto-ignition delay time using MG-L-PCA as a function of temperature for various system pressures, while using $q = 4$ (a) or $q = 5$ (b).

It should be stressed that the database used to train the model is the 1D laminar flame which was extended along the y -axis to initialize the computation (see Section 3.4.1). This 1D laminar flame is a syngas-air flame at $\phi = 0.88$; syngas and air have the same composition as in the 2D DNS. Grid consists in 2501 points with a spacing of $1.25 \cdot 10^{-5}$ m. Inlet velocity and temperature are set to 0.2 m/s and 300 K, respectively, and pressure is set to 101325 Pa.

During the computation, portions of the state-space not included in the original manifold are accessed due to the ability of the PCs to account for the flame-turbulence interactions. Moreover, only one flame was necessary to train the model and it provided the transported variables, and calculation of the scaling and centering coefficients for the full calculation. Comparison of Y_{H_2O} and Y_{HCO} fields are given in Figure 3.10, which shows a very good agreement between the solutions. Figures 3.11 and 3.12 show scatter plots of Y_{CO} , Y_H , $Y_{H_2O_2}$, Y_{HO_2} vs temperature for DNS and MG-PCA computations. Again, a very good agreement with DNS is observed. It appears that the model can naturally account for the turbulence-chemistry interactions of the system, thanks to the higher number of degrees of freedom available with respect to other methods such as ILDM or FPI and the appropriate selection for the key variables to be transported in the code. The fact that a single 1D laminar flame is sufficient to train the model for more complex flame-turbulence simulation is also appealing, as it proves that the applicability of the model is not limited to the system used to generate the database. Indeed, for a higher turbulence intensity and eddy size of the same order of magnitude of the flame, it is expected that a single flame might not be sufficient and that the effect of strain should be included in the manifold generation process.

3.6 Conclusions

PCA has demonstrated capability in identifying the low-dimensional manifold which can accurately describe a chemically-reacting system with a reduced number of optimal parameters. The MG-PCA approach provides a realistic application of PCA for combustion systems through the use of principal variables. A comparison with the classic PC-score approach indicated the strong potential of MG-PCA for the development of reduced-order combustion models of tailored accuracy. The present paper provides an easy and effective tool for the development of reduced models, whose implementation requires only minor modifications of existing CFD codes, and for which the accuracy of the model can be evaluated and tailored *a priori*. The main finding of the present paper can be summarized as follows:

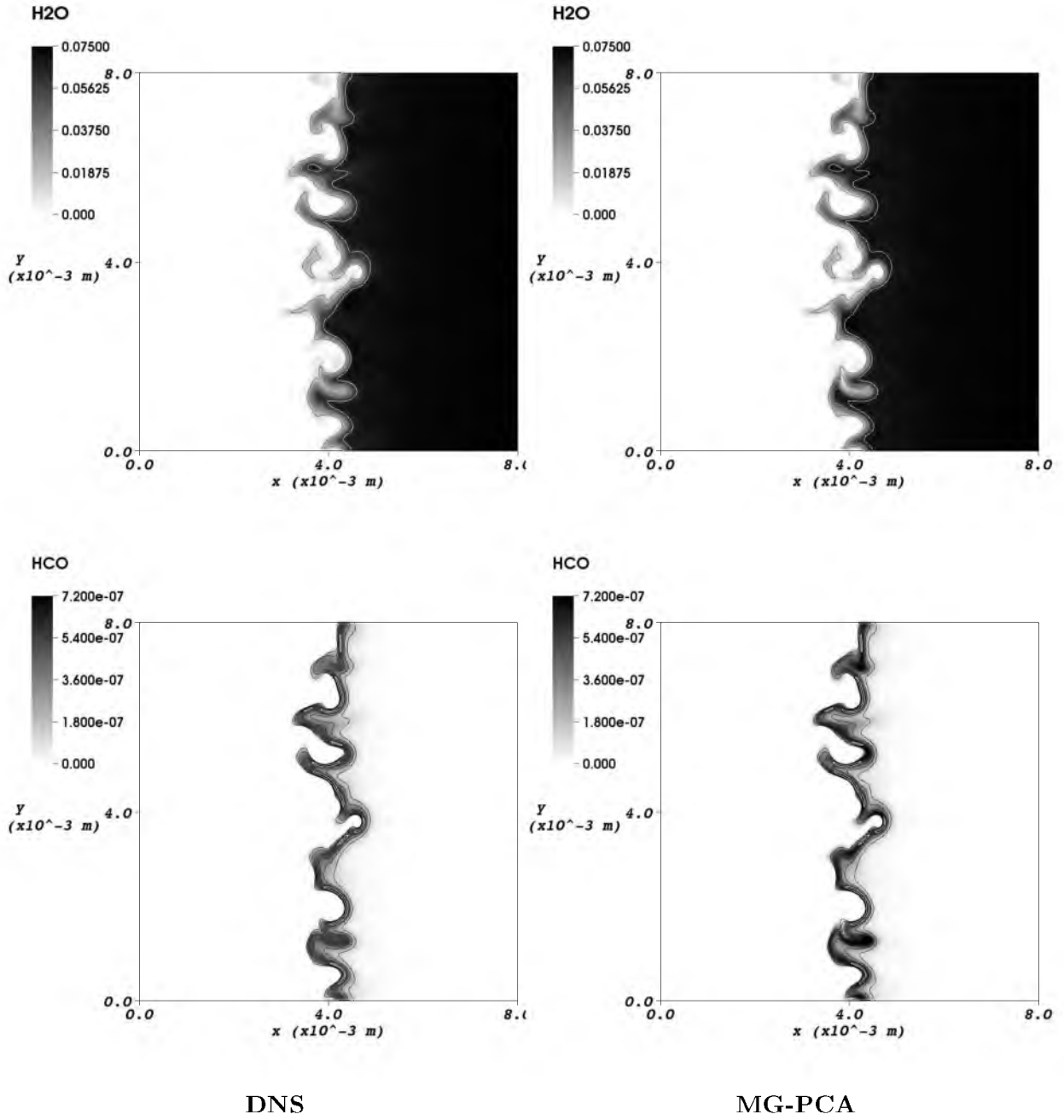


Figure 3.10: Fields of Y_{H_2O} (top) and Y_{HCO} (bottom) using DNS (left) and MG-PCA (right) $t = 7.11799 \times 10^{-4} \text{ s}$.

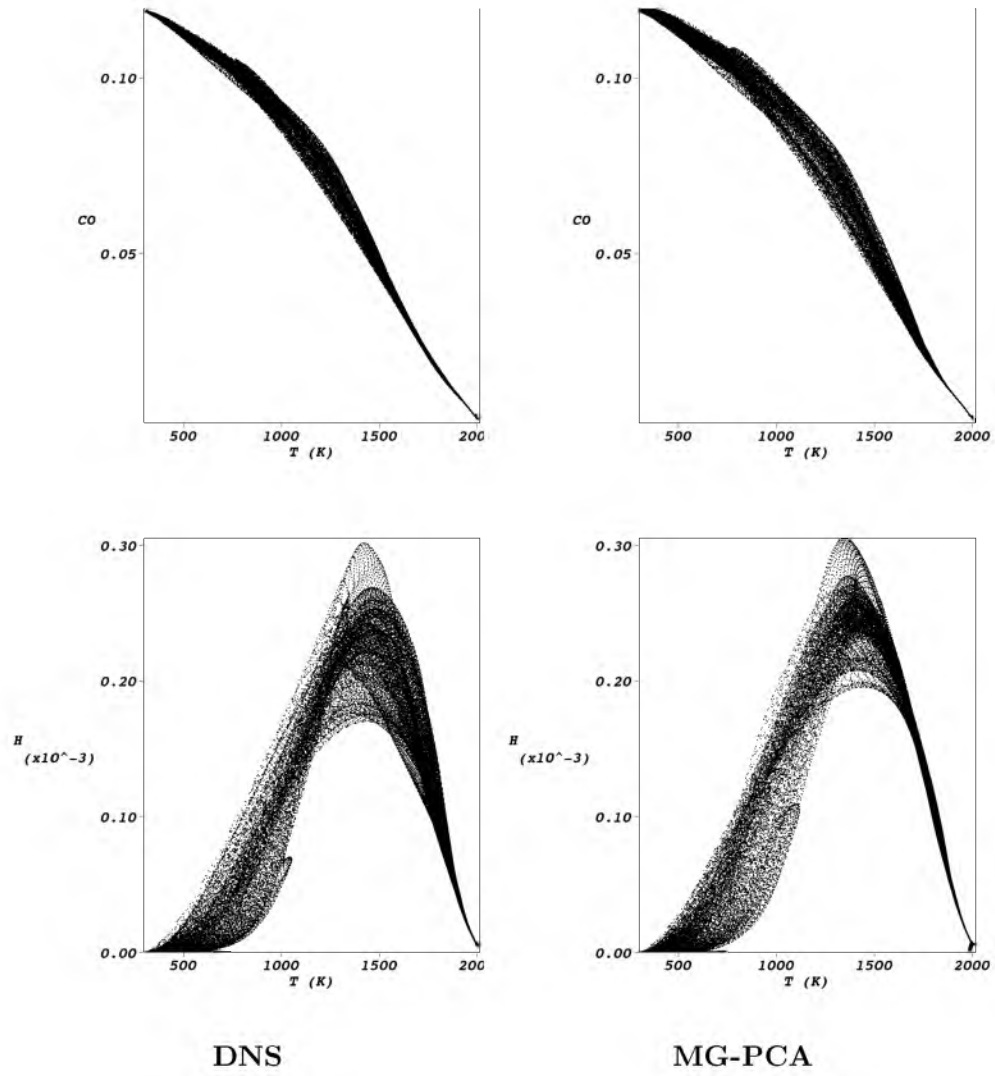


Figure 3.11: Scatter plot of (from top to bottom) Y_{CO} , Y_H vs temperature, using DNS (left) and MG-PCA (right) $t = 7.11799 \cdot 10^{-4}$ s.

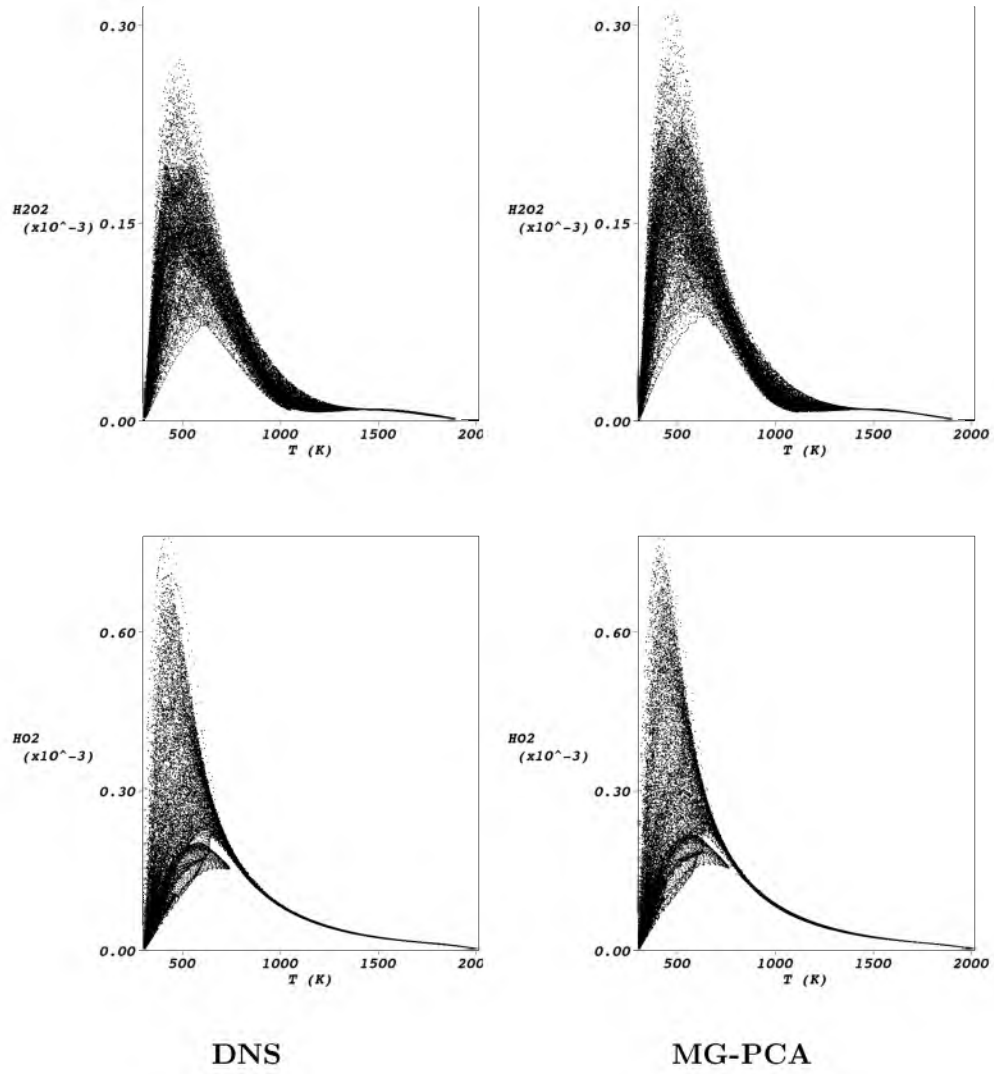


Figure 3.12: Scatter plot of (from top to bottom) $Y_{H_2O_2}$, Y_{HO_2} vs temperature, using DNS (left) and MG-PCA (right) $t = 7.11799 \cdot 10^{-4}$ s.

- The calculation of the \mathbf{B} matrix (Equation 3.13) using the actual \mathbf{Z}_q gives the MG-PCA methods an increased accuracy in reconstruction of nontransported state-space variables.
- The selection of the transported variables using the various principal variable selection methods greatly affects the reliability and accuracy of MG-PCA models, thus justifying the need for optimal selection techniques, as the ones outlined here. As far as the PV selection techniques are concerned, the B2 method was found to be the most robust and the one providing the best approximation of the state-space.
- Scaling methods play a major role in the identification of the optimal projection matrix \mathbf{A}_q and subsets of transported variables. In particular, it was shown that scaling methods other than the standard auto-scaling can also provide increased accuracy in reproducing state-space variables and principal variable source terms. Among them, auto, range, and pareto scaling provided the better results.
- The global MG-PCA approach can be effectively employed for simple fuels such as hydrogen [16], or slightly more complex fuels like syngas under unity Lewis number conditions. However, for larger mechanisms such as syngas or methane, and under differential diffusion conditions, the MG-L-PCA formulation must be employed, in order to achieve a significant reduction and to capture the nonlinear features of the actual manifold underlying the chemically reacting system.
- The flame-turbulence interaction computation demonstrated the ability of the MG-PCA method to perform an actual computation, it also demonstrated the good accuracy of the method. Moreover, the proposed model is not bounded to the original manifold used to train the model. Therefore, it naturally handles turbulence which is a clear advantage.

Future work will attempt to further automate the manifold generation procedure within CFD codes and to validate the overall approach on a broad range of combustion systems. Also, the coupling of the proposed MG-L-PCA clustering technique with a complete flow solver which should allow one to reduce arbitrarily the number of transported variables is considered.

3.7 Acknowledgments

This research was sponsored by the National Nuclear Security Administration under the Accelerating Development of Retrofittable CO_2 Capture Technologies through Predictivity

program through DOE Cooperative Agreement DE-NA0000740 and the Fonds National de la Recherche Scientifique, FRS-FNRS (Communauté Française de Belgique). Computing resources were provided by the IDRIS under the allocation 2009-i2009020164 made by GENCI (Grand Equipement National de Calcul Intensif).

CHAPTER 4

MODELLING COMBUSTION SYSTEMS WITH PRINCIPAL COMPONENT ANALYSIS TRANSPORT EQUATIONS

Benjamin J. Isaac, Jeremy N. Thornock, James Sutherland, Philip J. Smith, and
Alessandro Parente

Submitted to Journal of Computational Physics, March 2014

4.1 Abstract

Modelling the physics of combustion systems remains a challenge due to a large range of temporal and physical scales which are important in these systems. Detailed chemical kinetic mechanisms are used to describe the chemistry involved in the combustion process yielding highly coupled partial differential equations for each of the chemical species used in the mechanism. Recently, Principal Components Analysis (PCA) has shown promise in its ability to identify a low-dimensional manifold describing the reacting system. A PC-based model has been developed which may be well-suited for combustion problems; however, several challenging aspects of the model must be addressed. In this paper, the parameterization of state-space variables and PC-transport equation source terms are investigated. The ability to achieve highly accurate mapping through various nonlinear regression methods is shown. In addition, the effect of PCA-scaling on the ability to regress the surface is investigated. Finally, the present work demonstrates the capabilities of the model by solving a reduced system represented by several PC-transport equations for a perfectly stirred reactor (PSR) configuration and within a CFD solver simulating a synthesis-gas jet.

4.2 Introduction

The ability to accurately model a turbulent combustion system remains challenging due to the complex nature of combustion systems. A simple fuel such as CH_4 has been accurately described using 53 species and 325 chemical reactions [81]. More complex fuels require increasingly complex chemical mechanisms. Each resolved chemical species requires a conservation equation which is a coupled, highly nonlinear partial differential equation. Such systems are only possible to solve under very limited situations at this time due to computational costs. This issue leads to the need of a reduced model which can adequately describe the chemical reactions. Many methods attempt to reduce the complexity of the mechanism by splitting the system into slow and fast variables, using equilibrium assumptions for fast chemical processes, and occupying the computational resources on the more pertinent evolution of species within the reacting system [23, 34]. Indeed, in these complex combustion reaction mechanisms many of the species evolve at time-scales much smaller than the time-scales of interest, allowing for decoupling of fast and slow processes while maintaining accuracy. Low-dimensional manifolds exist in these systems which can describe the governing characteristics of the flames. Several models take advantage of this, including models such as the steady laminar flamelet model (SLFM) [66, 67, 70], flamelet-generated manifolds (FGM) [56, 89], or the flame prolongation of ildm (FPI) [27, 21, 22] to name a few. As a fundamental example, the steady laminar flamelet model uses the mixture fraction and mixture fraction variance to describe the flame as an ensemble of steady laminar diffusion flames undergoing various strain rates. In many cases, this provides a good representation of the entire system with a reduced number of variables.

Recently, principal component analysis (PCA) has been investigated for its use in combustion modelling. Several advantages of PCA include: its ability to identify orthogonal variables which are the best linear representation of the system; its ability to reduce in dimensionality requiring fewer coordinates; and the ability to do the analysis on canonical systems, such as the counter diffusion flame or empirical data-sets containing highly complex turbulent chemistry interaction. Parente et al. [61, 60] used PCA to identify the low-dimensional manifold in one-dimensional turbulence data, and experimental data. Biglari and Sutherland [5] and Pope [72] enhanced the capability of the PCA concept by combining the analysis with nonlinear regression, allowing a nonlinear mapping between state-space variables and the linear PCA basis. Mirgolbabaei and Echehki [50] extended this concept using artificial neural networks for the nonlinear mapping. In addition, several combustion models have been proposed based on the concepts from PCA. Sutherland and Parente [85]

derived transport equations for the principal components (PCs), and discussed the feasibility of a model where the PCs are used directly to construct state-space variables. Biglari and Sutherland [5] extended the concept of transporting the PCs by suggesting the nonlinear regression in order to increase the accuracy and reducibility of the model. Coussement et al. [16] and other groups [91] proposed transporting a reduced set of state-space variables and used the PC basis for reconstructing the variables which are not represented. Najafi-Yazdi et al. [52] used PCA to identify optimal progress variables to use the flamelet-generated manifold framework.

The present work seeks to advance the understanding and application of the PC-transport approach of Sutherland and Parente[85, 5] by first analyzing the effect of several scaling methods on the PC basis, and the resultant ability to regress the nonlinear state-space variables to the PC basis. Next, an analysis of nonlinear regression approaches is performed showing the advantages and disadvantages of the regression methods that have been suggested for the regression of the PC basis. Finally, an unsteady perfectly stirred reactor (PSR) calculation is shown using the PC-transport approach, and this is followed with a two-dimensional demonstration of the approach within a CFD solver.

Novel contributions of this work include a detailed comparison of regression methods for capturing the PC basis, including a discussion on the degree of nonlinearity of the state-space being represented by the PC basis, and finally, the first *run-time* demonstrations of the PC-transport approach within a numerical solver while using nonlinear regression. To the authors knowledge, all previous analyses of the original PC-transport concept using nonlinear regression have been done through *a priori* analysis on various data-sets [85, 50].

4.3 Data-sets

PCA-based models require a hi-fidelity data-set in order to derive the PC-basis, and fully parameterize the system. In general, a canonical reactor configuration which is appropriate for the system is used to generate a *training* data-set, which is then used in the PC analysis (see Section 4.4 for more detail) in order to construct the model. In the discussion that follows, several data-sets have been selected:

- A one-dimensional turbulence (ODT) data-set is used in Section 4.5.1 for an *a priori* analysis of the PC-transport model. The simulation is of a nonpremixed synthesis/air jet. For brevity, full detail concerning the simulation can be found in [29, 73]. The detailed reaction scheme for syngas [20] was used, the mechanism contains 11 chemical species (H_2 , O_2 , O , OH , H_2O , H , HO_2 , CO , CO_2 , HCO , N_2), and uses 21 chemical

reactions. The simulation is initialized with a temperature of $500K$, with air as the oxidizer ($0.7241 N_2$ and $0.2759 O_2$ by mass) and an N_2 diluted fuel stream containing $0.0078 H_2$, $0.5511 CO$, and $0.4411 N_2$ by mass. The ODT realizations are saved on a uniform grid of 672 grid points evenly spaced over a 0.01 m domain. The velocity field is initialized with a Reynolds numbers of 2500. The ODT data-set is particularly interesting because of the turbulence/chemistry interaction observed in the data, including physical effects such as extinction and re-ignition. In the analysis that follows, the capability of PCA in modelling these data is assessed.

- The unsteady solution to a perfectly stirred reactor (PSR) burning a stoichiometric mixture of syngas is used in Section 4.5.2.1. The data-set is generated by setting the inlet condition at $300K$ with a stoichiometric mixture of fuel and air, the vessel is initialized at equilibrium conditions (constant pressure and enthalpy), and multiple simulations are performed by varying the residence time in the vessel. All of the unsteady data for the various simulations is used collectively for the PCA analysis. The data-set is interesting in particular because it is trivial to solve, and does not contain any of the turbulent interactions of the previous data-set, yielding a smoother underlying manifold.
- Finally a laminar ODT solution is generated in order to give an ‘optimal’ manifold for the demonstration of the model within a CFD algorithm. The solution is optimal in the sense that absence of turbulent fluctuations leaves a smooth manifold which is easily regressed, when testing the model with nonlinear regression. The ODT simulation is performed with the same kinetic mechanism and diffusion model (mixture-average diffusion) of the ODT simulations discusses previously. The ODT code solves the laminar problem by suppressing the creation of eddies in the system, and allowing the system to diffuse. The inlet boundary condition is set by having the right half of the domain an inlet of air ($0.7241 N_2$ and $0.2759 O_2$ by mass) at 1 m/s and $300K$ and the left half as the fuel ($0.0078 H_2$, $0.5511 CO$, and $0.4411 N_2$ by mass) at 300 K with two different velocities. There is a transitional region between the fuel and oxidizer inlet where mixture fraction transitions smoothly from 1 to 0. The simulation is initialized with the solution to a highly strained counter-diffusion flame. The fuel stream velocities are initialized with either a higher velocity or a lower velocity, which tend to push the solution towards equilibrium or extinction, respectively.

Throughout the paper the coefficient of determination (R^2) and the normalized root mean square error (*nrms error*) are used as a means of quantifying the error produced through the application of the model:

$$R^2 = \frac{\sum_{i=1}^N (x_{predicted,i} - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (4.1)$$

$$nrms\ error = \frac{\sqrt{\sum_{i=1}^N (x_{predicted,i} - x_i)^2}}{max(\sigma(x_{predicted}, x))} \quad (4.2)$$

4.4 Theory

In this section, the basic concepts to PCA, the scaling used in PCA, and the various regression methods are presented.

4.4.1 Principal component analysis

A data-set consisting of n observations and Q independent variables is organized as an $n \times Q$ matrix (\mathbf{X}). The data \mathbf{X} is centered to zero by its corresponding means $\bar{\mathbf{X}}$, and scaled by the diagonal matrix, \mathbf{D} , containing the scaling value for each of the k variables:

$$\mathbf{X}^s = (\mathbf{X} - \bar{\mathbf{X}})\mathbf{D}^{-1} \quad (4.3)$$

In a PC analysis, the principal components (\mathbf{Z}) are identified by performing an eigenvalue decomposition of the covariance matrix of \mathbf{X}^s :

$$\frac{1}{Q-1} \mathbf{X}^{sT} \mathbf{X}^s = \mathbf{A}^{-1} \mathbf{L} \mathbf{A} \quad (4.4)$$

The eigenvector matrix \mathbf{A} (referred to here as a ‘basis matrix’) is then used to project the original state-space into PC space:

$$\mathbf{Z} = \mathbf{X}^s \mathbf{A} \quad (4.5)$$

Now given a subset of the basis matrix \mathbf{A} , denoted as \mathbf{A}_q and applying the previous equation, an approximation of the original centered and scaled state-space can be made using the following:

$$\mathbf{X}^s \approx \mathbf{Z}_q \mathbf{A}_q^T. \quad (4.6)$$

In the PC analysis, the largest eigenvalues correspond to the first columns of \mathbf{A} . This means the largest amount of variance in the original variables is described by the first PCs.

Accordingly, when one truncates the basis matrix (\mathbf{A}_q), the resultant approximation from Equation 4.6 may yield very accurate results, while representing the system with fewer variables.

In the work of Sutherland and Parente [85], a combustion model is proposed where conservation equations for the PCs are derived from the general species transport equation [71]:

$$\frac{\partial}{\partial t} (\rho Y_k) + \frac{\partial}{\partial x_i} (\rho u_i Y_k) = \frac{\partial}{\partial x_i} \left(\rho D_k \frac{\partial Y_k}{\partial x_i} \right) + R_k \quad (4.7)$$

One can easily derive the transport equations for the PCs (\mathbf{Z}_q) given the basis matrix \mathbf{A} , the scaling vector d_k , being the diagonal components of \mathbf{D} , and the centering vector \bar{Y}_k :

$$\frac{\partial}{\partial t} (\rho Z_q) + \frac{\partial}{\partial x_i} (\rho u_i Z_q) = \frac{\partial}{\partial x_i} \left(D_{Z_q} \frac{\partial}{\partial x_i} (Z_q) \right) + s_{Z_q} \quad (4.8)$$

$$s_{Z_q} = \frac{1}{\rho} \sum_{k=1}^Q \frac{R_k}{d_k} A_{kq} \quad (4.9)$$

According to the proposed formulation, one can theoretically utilize PCA with its inherent advantages. These advantages include: the ability to represent the system with a reduced number of variables; the option to include a predetermined amount of reconstruction error (dependent on q , the number of retained PCs), and possibly a reduction in stiffness if the selected PCs are highly weighted with major species.

The scaling matrix \mathbf{D} from Equation 4.3 plays a crucial role in PCA. Without scaling, it may be difficult to compute the correlation structure of variables with different magnitudes. The following scaling methods were adopted for this study [33, 63]:

- *auto scaling* (std), uses the standard deviation s_k . Auto scaling leaves all columns of $\tilde{\mathbf{X}}$ with a standard deviation of one, and now the data are analyzed on the basis of correlations instead of covariances, $d_k = s_k$.
- *range scaling* (range), uses the difference between the minimum and the maximum variable value, $d_k = \max(X_k - \bar{X}_k) - \min(X_k - \bar{X}_k)$.
- *pareto scaling* (pareto), adopts the square root of the standard deviation as scaling factor, $d_k = \sqrt{s_k}$.
- *variable stability scaling* (vast), gives an emphasis to variables which do not show strong variation, by using the product between the standard deviation and the coefficient of variation, $d_k = s_k \frac{s_k}{\bar{X}_k}$.

- *level scaling* (level), uses the mean value of the variables $d_k = \bar{X}_k$.

As will be shown, the scaling of the data-set effects the accuracy of the approximation made in Equation 4.6. The weakest issue with PCA-based models, for use in combustion systems, is that the linear model is attempting to model a highly nonlinear system. In order to take full advantage of the PC analysis, Biglari and Sutherland [5] suggests applying a nonlinear mapping to the linear underlying surface by using nonlinear regression. It has been shown [5, 50, 72] that the nonlinear regression allows one to fully utilize the underlying manifold identified by the principal component analysis. It is important to note that the linear basis derived from the PC analysis is critical as it allows for the derivation of simple transport equations; however, by using nonlinear functions within this basis, the model is allowed to capture the nonlinearities which are always present in combustion systems.

4.4.2 Regression models

In this study, nonlinear regression is used to model the highly nonlinear state-space variables as a function of the principal components (\mathbf{Z}). In place of Equation 4.6, now the various state-space variables and PC source terms ($\mathbf{s}\mathbf{z}$) are mapped to the PC basis using the nonlinear regression function f_{Φ} :

$$\Phi \approx f_{\Phi}(\mathbf{Z}_q) \quad (4.10)$$

where Φ represents the state-space variables, or in terms of regression, the dependent variables (i.e. Y_i , T , ρ , and, $\mathbf{s}\mathbf{z}$).

Until now, two nonlinear regression methods have been applied to mapping Φ to \mathbf{Z} . In the work of Biglari and Sutherland [5] and Pope [72], multivariate adaptive regression splines are used. In the work of Mirgolbabaei and Echehki [50], artificial neural networks are investigated. Here, in addition to previously used regression techniques, several other methods are investigated, including support vector regression, and gaussian process regression. In summary, the following regression techniques are investigated:

- *Linear Regression Model* (LIN)

The linear model applied in multiple dimensions is of the form:

$$\Phi = \mathbf{Z}\mathbf{a} + \mathbf{v} \quad (4.11)$$

where \mathbf{a} is the regression coefficient vector and \mathbf{v} is the intercept vector [11].

- *Multivariate Adaptive Regression Splines* (MARS)

Multivariate adaptive regression splines use the concept of building up the model from

product spline basis functions. This model creates a number of basis functions, and automatically determines knot location and implements splines at knot boundaries. The model is of the form:

$$\Phi = \sum_{m=1}^M a_m B_m(\mathbf{Z}). \quad (4.12)$$

where B_m are the basis functions and a_m are the expansion coefficients [24].

- *Artificial Neural Networks (ANN)*

Artificial neural networks uses the concept of networking various layers of estimation resulting in a highly accurate output layer. Following the theory of Pao [57], the model works as follows: first, t hidden networks (NET_t) are calculated as a weighted (w_t) sum of the training data inputs ($k_i = [\mathbf{Z}, \Phi]$):

$$NET_t = \sum_{i=1}^N w_{ti} k_i + b_i. \quad (4.13)$$

A sigmoid transfer function is then used to generate an output for the network:

$$Z_t = [1 + \exp(-NET_t)]^{-1}. \quad (4.14)$$

Next, the output networks are calculated:

$$NET = \sum_{t=1}^h \nu_t Z_t + b_o \quad (4.15)$$

Again, the network is scaled and a prediction of Φ is then given:

$$\Phi = [1 + \exp(-NET)]^{-1}. \quad (4.16)$$

In the present study, one hidden layer with 20 neurons is used with one neuron in the output layer.

- *Support Vector Regression (SVR)*

Support vector regression is a subset of support vector machines (SVM). The idea behind SVR is again to create a model which predicts \mathbf{sZ} given \mathbf{Z} using learning machines which implement the structural risk minimization inductive principle. The basic model form is

$$\Phi = \sum_{i=1}^N (\alpha_i^* - \alpha_i) K(\mathbf{Z}_0, \mathbf{Z}_i) \quad (4.17)$$

where α_i^* and α_i are Lagrange multipliers, and $K(\mathbf{Z}_0, \mathbf{Z}_i)$ is the kernel operator [82]. In the current study, a radial-based kernel was used and the optimum kernel hyperparameter as well as the insensitive-loss function were determined by doing various calculations over a range of input parameters.

- *Gaussian Process Regression* (GPR)

Gaussian process regression is founded on the idea that dependent variables can be described by a gaussian distribution [53, 75]:

$$\Phi \sim N(0, \mathbf{K}(\mathbf{Z}, \mathbf{Z}) + \sigma_n^2 \mathbf{I}) \quad (4.18)$$

Here \mathbf{Z} is the data matrix containing all sample points in PC space; $\mathbf{K}(\mathbf{Z}, \mathbf{Z})$ is the kernel function for \mathbf{Z} ; in the current study, the gaussian kernel is used:

$$\mathbf{K}(\mathbf{Z}_p, \mathbf{Z}_q) = \sigma_s^2 \exp\left(-\frac{1}{2}(\mathbf{Z}_p, \mathbf{Z}_q)^T \mathbf{W}(\mathbf{Z}_p, \mathbf{Z}_q)\right). \quad (4.19)$$

Given query points \mathbf{Z}_* it can be shown that a prediction Φ_* can be made using the following formula:

$$\Phi_* = \mathbf{K}_*^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \Phi \quad (4.20)$$

where $\mathbf{K}_* = \mathbf{K}(\mathbf{Z}, \mathbf{Z}_*)$ and $\mathbf{K} = \mathbf{K}(\mathbf{Z}, \mathbf{Z})$. The initial guess for the kernel's hyper-parameters: the characteristic length scale, and signal variance, were one. A gradient-based marginal likelihood optimization was used to determine these values.

4.5 Results and discussion

In the current section, the feasibility of the model is assessed on an ODT nonpremixed data-set. Both scaling and regression methods are tested on the data. Based on the conclusions from the *a priori* study the model is then tested in a PSR, and the initial results to 2D laminar simulation are presented utilizing the presented approach.

4.5.1 A priori model evaluation

The effect of the various scaling methods and the optimal regression method are now tested in order to assess the feasibility of the model on a combustion data-set. The current section analyzes the ODT data-set discussed in Section 4.3.

4.5.1.1 Scaling

The chemical species mass fractions \mathbf{X} (where $Q = 11$) are now tested for the various scaling methods discussed in Section 4.4.1, using Equation 4.6. Figure 4.1 shows the *nrms error* for the reconstruction of several major species on the left (CO , CO_2 , H_2O), and several radical species on the right (HCO , HO_2 , OH) while varying q . It is obvious from Figure 4.1 that pareto scaling had a distinct advantage for the major species, mostly due to the fact that the scaling is $1/\sigma^2$. In general, most of the scaling methods have trouble with the smaller variables, such as the radicals, with level scaling appearing to give

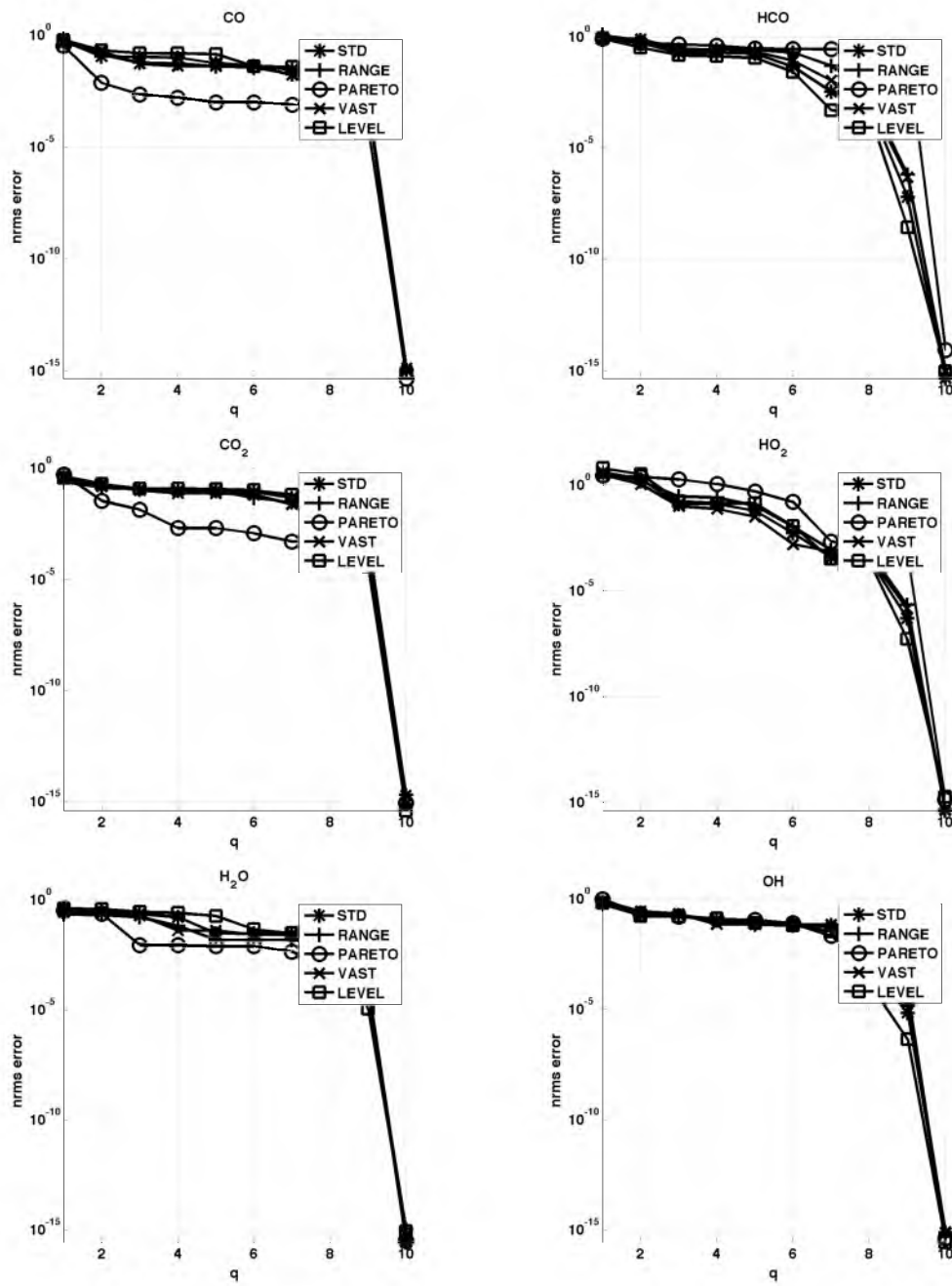


Figure 4.1: *Nrms* error values for several major species (left), and minor species (right), while varying q , the number of PCs, and the scaling method.

the best results for the current case. The implications of the various scaling strategies are profound. In the case of pareto scaling, the weighting is attractive as it gives more weight and importance to major variables. The scaling effects are important as well when looking at the ability of the reduced model when computing the reaction source terms \mathbf{s}_Z .

The calculation of the chemical species reaction rates is done using Cantera [28]. The reaction rates are now computed from the original \mathbf{X} and the approximation, using Equation 4.9. Figure 4.2 shows the *nrms error* for s_{Z_1} for the various scaling methods while varying q . An *nrms error* lower than 10^{-2} is not achieved until $q = 8$ for all of the scaling methods. This indicates that the error from Equation 4.6 is propagating into the calculation of \mathbf{s}_Z .

Due to the linear nature of the PC-based model, a large value for q is required to accurately recover the highly nonlinear reaction rates. With differential diffusion, enthalpy and elemental mass fractions are not constant, yielding 11 degrees of freedom for the ODT data-set. With $q = 8$, only a minor reduction is achieved. An alternative to the direct reconstruction of \mathbf{X} is to use nonlinear regression functions, which can be used to map the nonlinear reaction rates or nonlinear species concentrations to the lower dimensional representation given by the PCs.

4.5.1.2 Regression

In order to map the highly nonlinear reaction rate surface (dependent variables) to PC space (independent variables) it is useful to understand how nonlinear the reaction rates and other state-space variables are with respect to the underlying manifold represented by the principal components. A simple way to do this in multiple dimensions is to divide the independent variable space onto a coarse grid, and assess locally the variation of dependent variables within a local section of the independent variable space. Locally, if the dependent variable has a large variation, then the ability to regress the dependent variable locally will be more difficult because of the nonlinear nature or even local scatter in the data. The following equation is used to calculate the locally normalized variance for the i^{th} coarse grid cell (χ_{Φ}^i):

$$\chi_{\Phi}^i = \frac{\nu(\Phi(\mathbf{Z}_{\mathbf{q}}^i))}{\nu(\Phi(\mathbf{Z}_{\mathbf{q}}))} \quad (4.21)$$

where $\nu(x) = \langle (x - \langle x \rangle)^2 \rangle$ is the variance function which is calculated on the observations within the i^{th} coarse grid cell ($\Phi(\mathbf{Z}_{\mathbf{q}}^i)$) or for all observations ($\Phi(\mathbf{Z}_{\mathbf{q}})$). Now, summing over all course grid cells in PC space, we obtain the overall manifold nonlinearity for dependent variable Φ :

$$\chi_{\Phi} = \sum_{i=1}^c \chi_{\Phi}^i \quad (4.22)$$

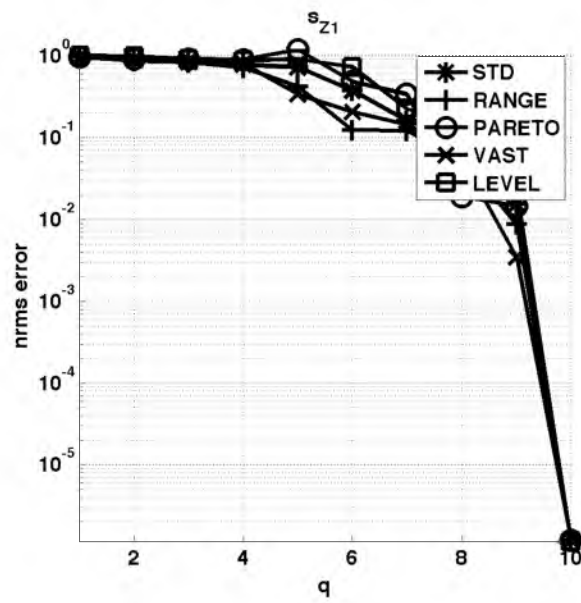


Figure 4.2: *Nrms error* values for s_{Z_1} while varying q , the number of PCs, and the scaling method.

Table 4.1 shows the manifold nonlinearity calculation for various dependent variables while changing the scaling methods. It is clear from the analysis that some scaling methods have distinct advantages for several of the dependent variables. In particular, pareto scaling has an advantage when comparing several major species (O_2 , CO , CO_2 , and N_2), temperature, and density, with a weaker performance for some of the radical species (OH , H). All methods show the regression for s_{Z_1} is challenging; however, the regression for s_{Z_2} appears promising with pareto scaling.

Given the results for both the state-space reconstruction, and the manifold nonlinearity, it is clear that the pareto scaling method has some unique advantages for this particular data-set dealing with syngas combustion. With this observation in mind, the various regression models are now tested with the pareto scaling method. The nonlinear regression analysis is done using a combination of computing software packages including the statistical computing software R [74], and MATLAB [47]. The R code implementations for LIN, MARS, ANN, and SVR were used. For GPR, the MATLAB toolbox gpml [75] was employed. The analysis is done on $n = 5000$ sample points evenly distributed over \mathbf{Z} space. The analysis is done using $q = 2$ and 3.

Table 4.2 shows the regression results for s_{Z_1} as a function of \mathbf{Z} , with $q = 2$ and $q = 3$. As expected, the linear regression method has difficulty mapping the highly nonlinear dependent variables. Complex methods also struggle with the mapping while $q = 2$. When moving to $q = 3$, the later 3 methods are beginning to show higher accuracy. In this particular case, GPR produces the most accurate reconstruction. The approximation shows a vast improvement especially if compared with the results of the direct computation (Equation 4.9), with the same level of accuracy being achieved with $q = 8$.

4.5.1.3 Subset PCA

In the work of Mirgolbabaei and Echehki [50], the PCA analysis is done on a subset of species in order to recover sufficiently accurate source terms. This has the benefit of removing certain species which may be contributing highly nonlinear source terms to \mathbf{s}_{Z_q} . The drawback to doing this is that there is no guarantee that the underlying manifold computed from the subset will be able to adequately predict the species removed from the analysis. In the current study, the retained species are selected by choosing variables which tend to pertain to the slower chemical time-scales of the system, such as the major species. The following subset of species were selected for the present analysis: H_2 , O_2 , H_2O , CO , and CO_2 .

With the selected subset of species, the PCA analysis is repeated, again with pareto

Table 4.1: Manifold nonlinearity (χ_{Φ}) for state-space variables, Φ while using different scaling methods.

χ_{Φ}	std	range	pareto	vast	level
H_2	5.7	11.4	10.8	12.3	3.5
O_2	4.0	1.9	0.3	0.7	4.9
O	12.6	11.8	17.2	28.8	7.5
OH	16.6	17.3	21.5	41.5	6.8
H_2O	6.1	5.1	4.9	5.3	7.0
H	14.6	22.3	30.0	46.1	5.2
HO_2	7.1	9.6	6.2	3.3	7.2
CO	2.4	1.3	0.1	1.8	1.7
CO_2	5.0	5.0	0.8	3.0	6.2
HCO	6.9	14.6	18.1	29.4	2.5
N_2	1.7	0.7	0.1	0.7	1.4
T	7.0	6.5	2.0	4.0	9.2
ρ	7.8	6.9	2.5	5.0	9.6
s_{Z_1}	256.5	292.2	300.5	404.0	210.1
s_{Z_2}	150.0	172.7	25.8	143.7	95.9

Table 4.2: *Nrms error* and R^2 statistics for the prediction of s_{Z_1} while using pareto scaling and $q = 2$ or $q = 3$.

Method	<i>nrms error</i> ($q = 2$)	R^2 ($q = 2$)	<i>nrms error</i> ($q = 3$)	R^2 ($q = 3$)
LIN	0.99	0.02	0.67	0.55
MARS	0.30	0.91	0.26	0.93
ANN	0.22	0.95	0.20	0.96
SVR	0.23	0.95	0.19	0.97
GPR	0.22	0.95	0.18	0.97

scaling. Figure 4.3 shows the scree plot [33], which gives the percentage of variance accounted for while selecting q PCs. The figure compares the full PCA version using 11 variables and the subset PCA using 5. It is clear that the PCA based on the subset of variables represents the variation in the system with fewer variables.

Table 4.3 shows the error statistics for the entire set of Φ while using gaussian process regression and pareto scaling. It is interesting to note even though several of these variables were not included in the analysis, the PCA basis computed from the major species in combination with the nonlinear regression is sufficient for mapping these highly nonlinear minor species.

One of the biggest advantages behind using PCA for combustion modelling is PCA's ability in deriving orthogonal variables which best parameterize the reaction system, through a linear representation. It is useful to relate the PCs to physical variables. In the current study, the eigenvector weights reveal interesting details into what the PCs physically mean. Table 4.4 shows the basis matrix weights from the PCA analysis on the major species. The weights from the first PC have large positive values for carbon containing variables (CO , CO_2), and a large negative value on the oxidizer (O_2). This appears to be very similar in nature to Bilger's mixture fraction [6], ξ . Figure 4.4 shows a plot of Z_1 against ξ ; the plot shows that Z_1 is clearly correlated with ξ . The weights for Z_2 show positive correlations for H_2 , O_2 , and CO , with negative correlations for H_2O and CO_2 . These weights appear to be related to the extent of reaction, where reactants have negative stoichiometric coefficients, and products have positive reaction coefficients. With a larger initial mass-based concentration of CO (compared with H_2), a large amount of CO_2 is produced, and a much smaller amount of H_2 is present leading to a smaller positive weight on H_2 and smaller negative weight for the product H_2O .

It is interesting to point out that without any prior understanding or assumptions on the combustion systems, the PC analysis has automatically identified two important variables which are often used to characterize combustion systems.

It is evident that the linear PC model in conjunction with a nonlinear regression has the potential of delivering very accurate state-space variables as well as reaction rates for a given system of interest. To the authors knowledge, no actual computation of the PC-transport approach exist in the literature. The following section gives two simplified demonstrations of the model within a numerical solver.

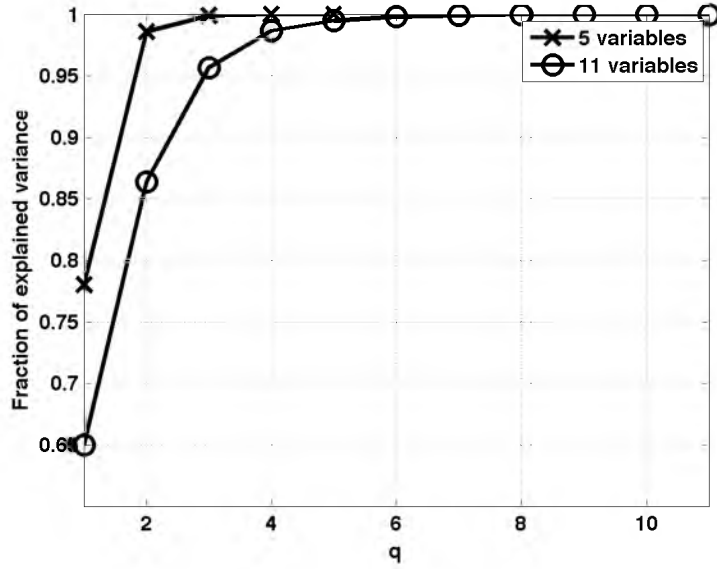


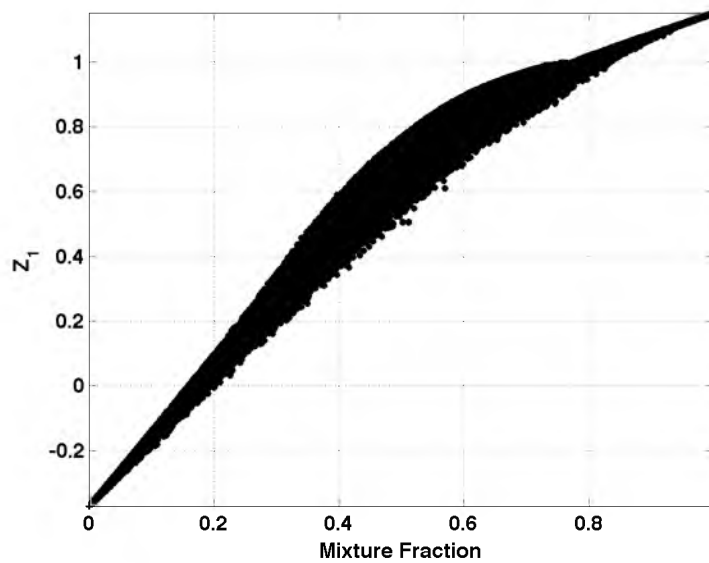
Figure 4.3: Scree plot from the eigenvalue matrix, showing the fraction of explained variance (y-axis) as a function of the number of PCs (q) for the system containing a subset of the original species ('x' markers), and the full system ('o' markers).

Table 4.3: *nrms error* and R^2 statistics for the prediction of Φ while using pareto scaling and $q = 2$.

Φ	<i>nrms error</i>	R^2
H_2	0.05	0.997
O_2	0.04	0.999
O	0.06	0.996
OH	0.07	0.995
H_2O	0.06	0.997
H	0.05	0.997
HO_2	0.17	0.969
CO	0.05	0.998
CO_2	0.05	0.997
HCO	0.03	0.999
N_2	0.04	0.998
T	0.04	0.998
ρ	0.04	0.999
s_{Z_1}	0.22	0.949
s_{Z_2}	0.16	0.974

Table 4.4: Eigenvector matrix, A , from the PC analysis.

species weight	Z_1	Z_2	Z_3	Z_4	Z_5
H_2	0.047	0.117	-0.302	0.900	0.288
O_2	-0.627	0.119	-0.034	-0.230	0.734
H_2O	0.176	-0.186	0.895	0.222	0.292
CO	0.624	0.656	-0.040	-0.243	0.348
CO_2	0.431	-0.713	-0.325	-1.124	0.414

**Figure 4.4:** A scatter plot of mixture fraction (x-axis) versus Z_1 (y-axis), illustrating the correlation between the variables.

4.5.2 A posteriori model evaluation

In order to demonstrate the approach a perfectly stirred reactor and a 2D jet flame are investigated. First, the proposed method is demonstrated in a PSR, comparing the calculations using the full set of equations to the standard PC-transport approach, and the PC-transport approach using nonlinear regression. Second, a 2D syngas jet is calculated using the PC-transport approach with nonlinear regression. The computation is done in ARCHES, a CFD software developed during the last decade through awards granted between the University of Utah and the Department of Energy [77, 76].

4.5.2.1 Perfectly stirred reactor

An implementation for the perfectly stirred reactor was made using MATLAB. The following governing equations were implemented and solved using the ccode toolbox in MATLAB [14]:

$$\frac{d\rho Y_i}{dt} = \frac{\rho}{\tau} Y_i^0 - \frac{\rho}{\tau} Y_i + R_i W_{s,i} \quad (4.23)$$

where Y_i and R_i are the i^{th} species mass fraction and molar reaction rate ($kmole/m^3/s$), τ (seconds) is a constant representing the residence time through the reactor, $W_{s,i}$ is the i^{th} species molecular mass, and ρ is the density (kg/m^3). The temporal solution to the equations are solved using the Newton nonlinear solver, and the BDF multistep method. The problem is initially solved using a stoichiometric mixture of syngas-air using the same mechanism which was used for the ODT data-set ([20]), where the mechanism includes 11 chemical species and 21 reactions. The inlet conditions for the reactor (Y_i^0) are set at an equivalence ratio of 1 with a temperature of 300K. The initial conditions for the reactor (Y_i) are set at the equilibrium conditions of the inlet and the system is run until a steady-state solution is reached. The PSR is modelled assuming constant volume, residence time, and pressure. Multiple simulations are performed by varying the residence time and saving the temporal solution until steady-state is reached. The PCA process described in Section 4.4.1 is then applied to the data to create the basis matrix \mathbf{A}_q , and the regression functions f_Φ for the state-space variables, Φ . The approach is then tested with various values of τ , which were not used when creating the data-set.

The regression of Φ is carried out using $q = 2$ resulting in R^2 of 0.9995 or higher for all variables including s_{Z_q} . The simulations are then performed with 2 transport equations instead of 11, yielding a significant reduction. Figures 4.5a-4.10b show the temperature and species mass fractions of the system. The markers show the steady-state solution for a given τ using the PC-transport model. The underlying solid-lines in the figures show the

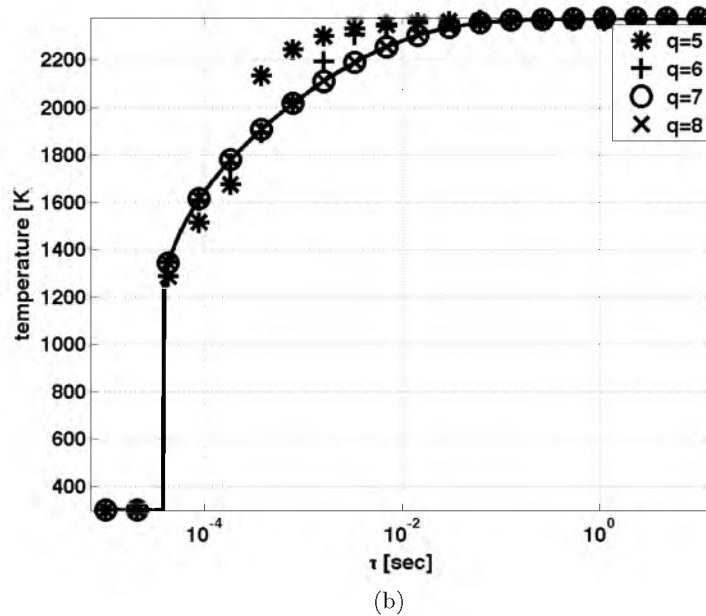
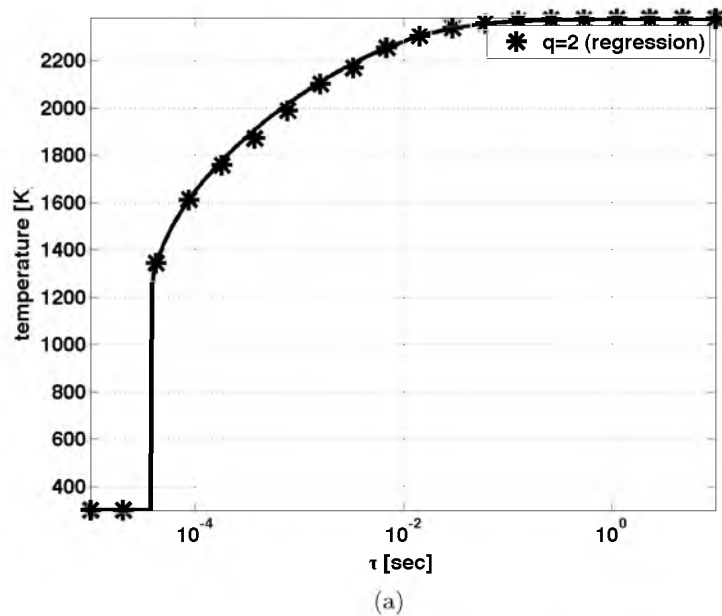
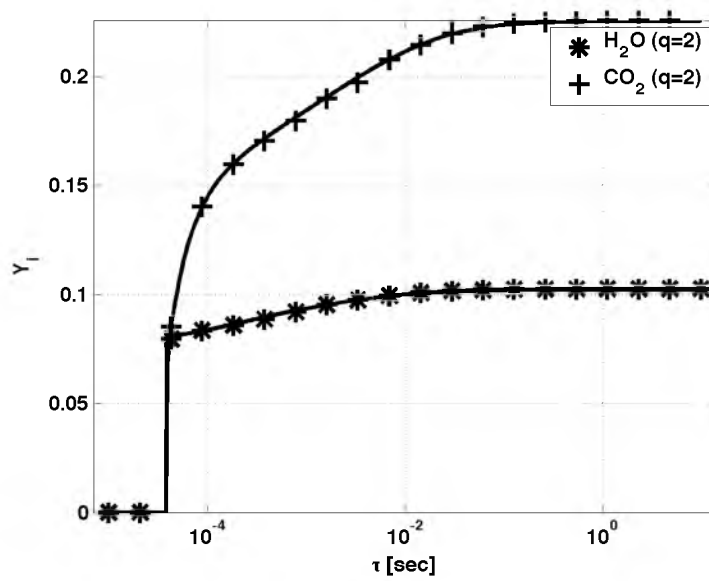
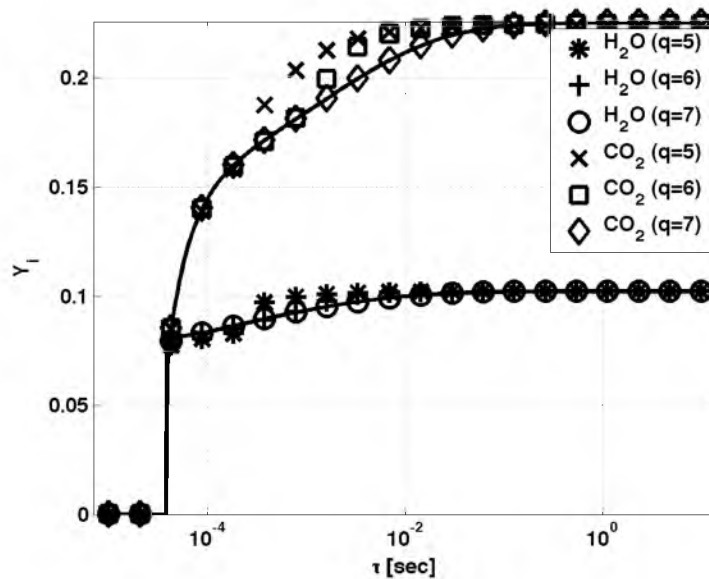


Figure 4.5: PSR temperature as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q .



(a)



(b)

Figure 4.6: Major species products as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q .

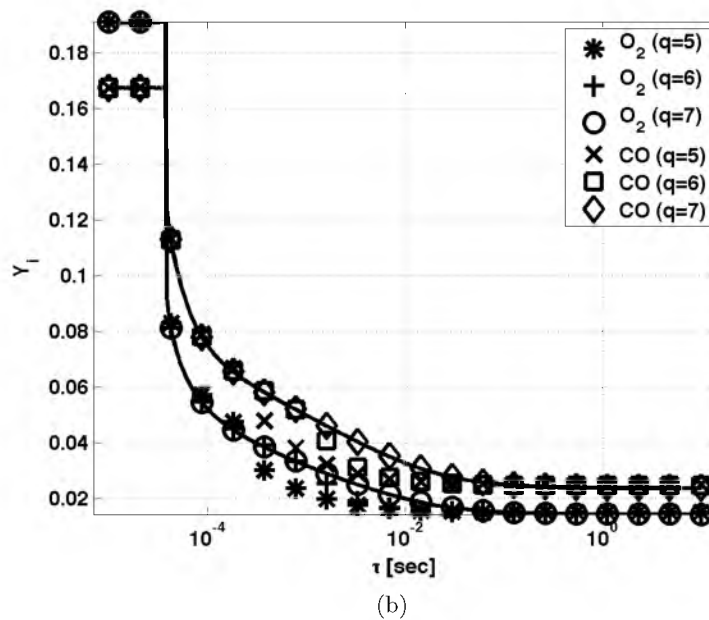
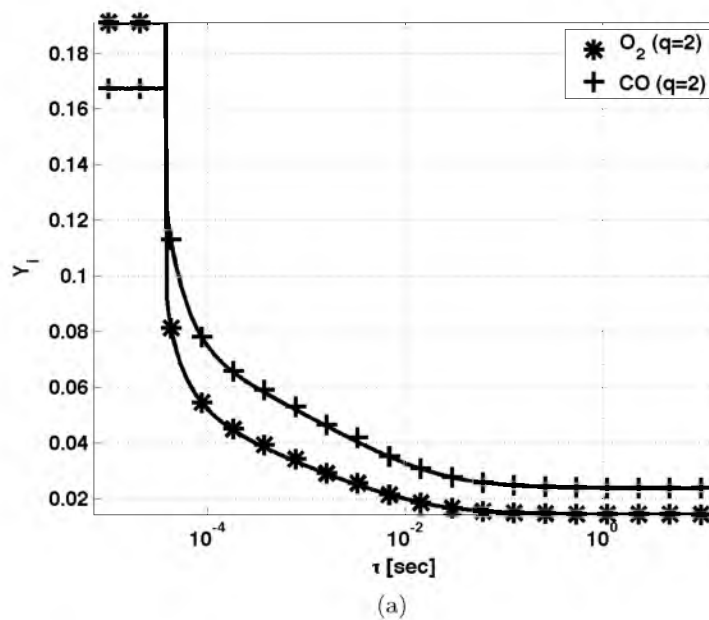


Figure 4.7: Major species reactants as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q .

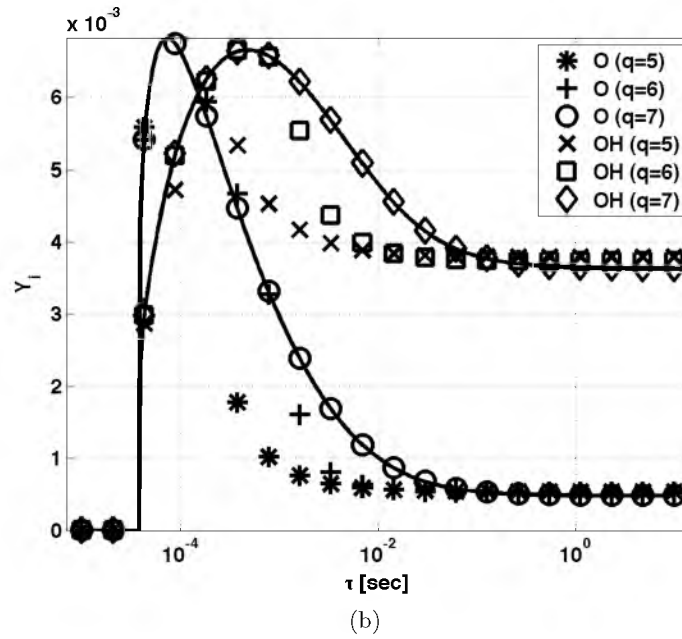
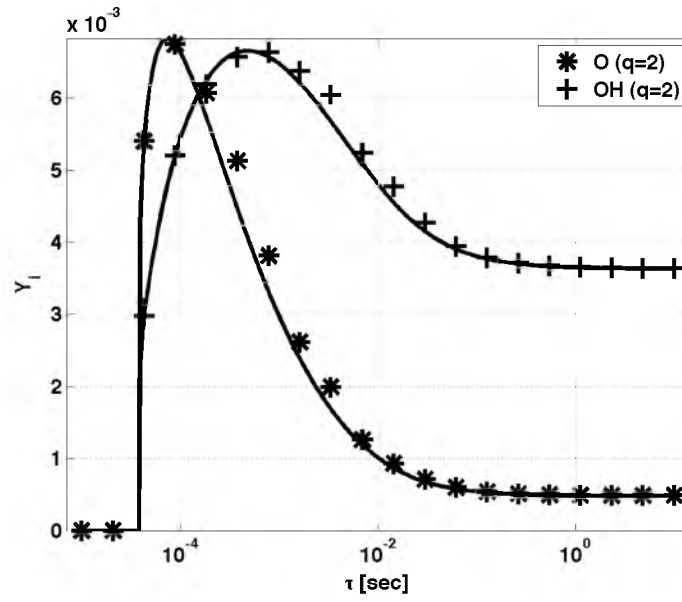
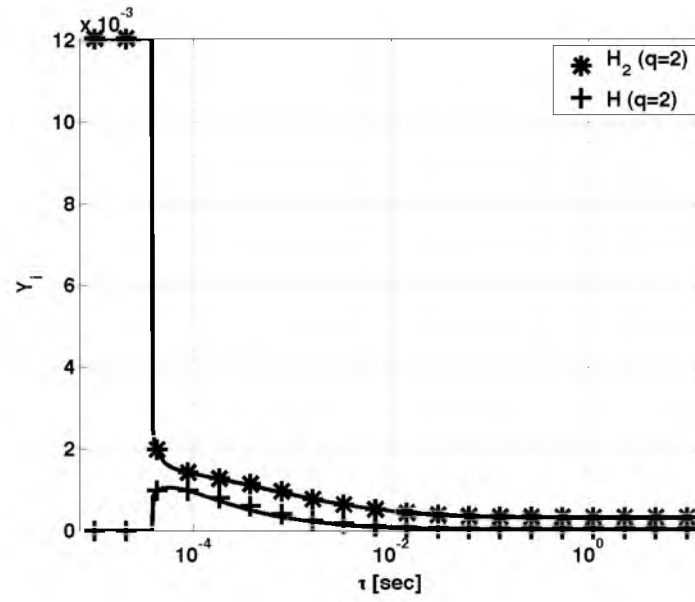
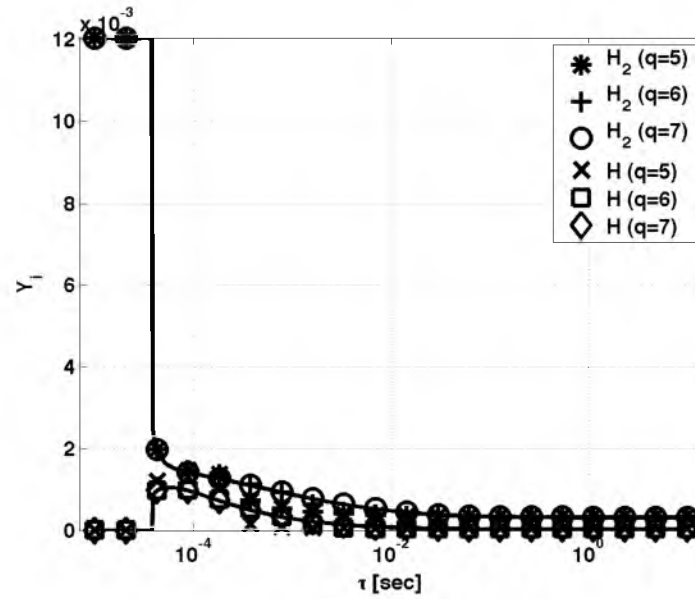


Figure 4.8: Minor species as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q .

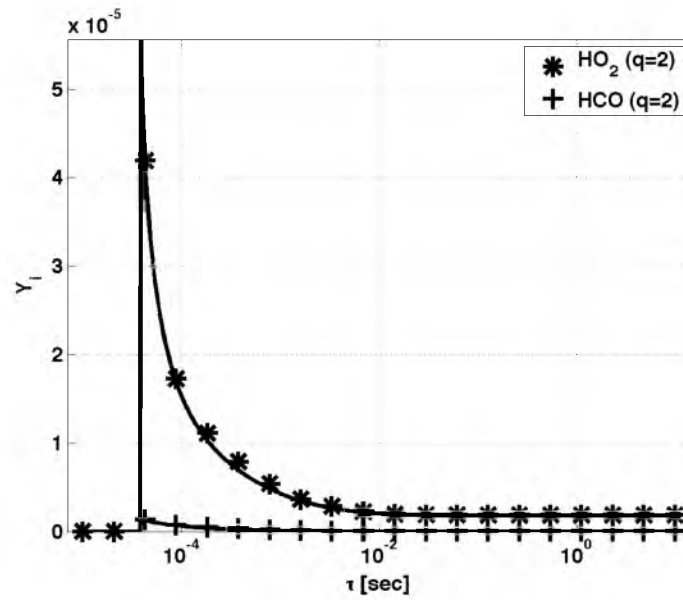


(a)

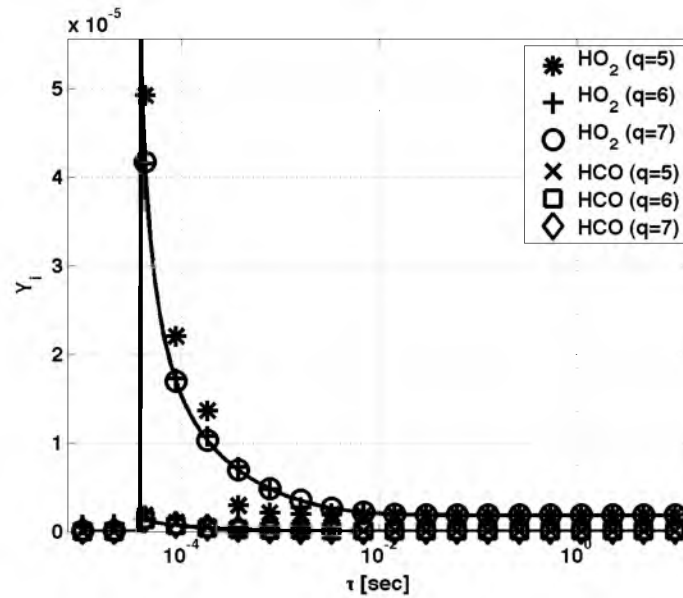


(b)

Figure 4.9: Minor species as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q .



(a)



(b)

Figure 4.10: Minor species as a function of the residence time, with the solid-line representing the full solution. The markers represent the results for the model with GPR regression (a) using $q = 2$ PCs, and the standard model without regression (b) while varying q .

full solution calculated over a range of residence times. The top plot (a) shows the results of the model using GPR for the nonlinear mapping with $q = 2$, and the results on the bottom (b) show the standard model without the regression step while varying q . The results show remarkable accuracy for the model with regression over the range of residence times for the predicted temperatures, and both major and minor species. A similar degree of accuracy is not observed in the model without regression until $q = 7$. In the current system, constant enthalpy and elemental mass is observed yielding 7 degrees of freedom, which would imply virtually no reduction due to the degrees of freedom.

4.5.2.2 Syngas jet flame

In order to provide a demonstration of the approach within a CFD solver, a combustion data-set with several important features is needed:

- *Representative Data-set*

The data-set used to train the PC-transport model must be representative of the system of interest. This clearly implies that a canonical reactor of lesser computational costs needs to be used to develop the data-set. It has been demonstrated [5] that the low-dimensional manifolds may in fact be invariant under certain conditions, allowing a system to be modeled using PCA calculated from a similar combustion case. In addition, the model is describing the temporal evolution of the species in the system; accordingly, the data-set must contain the temporally evolving profiles of the species.

- *Source Terms*

In order to use the approach in combination with nonlinear regression for \mathbf{s}_Z , the chemical species source-terms must be available. Due to this, the use of experimental data-sets would be unlikely, and numerical data-sets which employ detailed kinetic mechanisms are more appropriate.

- *Manifold Limits*

Again, because all of the state-space variables (including ρ) are tabulated before hand, \mathbf{Z} space must span all realizable values. In the context of nonpremixed combustion, the limits of the manifold may likely exist when the system exhibits mixing with no reactions occurring, and the equilibrium solution.

- *Accurate Regression*

In order to preserve mass, and stay within the limits of the manifold, the regression of the source-terms, in particular, needs to be accurate.

For a demonstration of the approach within a CFD solver, a data-set meeting the aforementioned criterion is needed. Accordingly, the laminar data-set mentioned in Section 4.3 was developed.

The PC analysis of the laminar data-set was done using the same subset of chemical species in Section 4.5.1.3 (H_2 , O_2 , H_2O , CO , and CO_2). Table 4.5 shows the results for the gaussian process regression of Φ for the laminar system. The weights of \mathbf{A}_q are given in Table 4.6, and the correlation between the Z_1 and mixture fraction is shown in Figure 4.11. Again, it is observed that Z_1 and Z_2 are similar to mixture fraction and extent of reaction. It is important to note that even though high correlation between Z_1 and mixture fraction is observed, s_{Z_1} is nonzero. However, as one would guess, s_{Z_1} is much smaller than s_{Z_2} (2 orders of magnitude, when comparing means).

With an accurate reconstruction of Φ , a 2D syngas jet flame is carried out solving for 2 transport equations ($q = 2$). The jet's inlet boundary consists of a partially premixed mixture of syngas and air with a mixture fraction of 0.79, at equilibrium conditions ($T \approx 1025$ K). The jet has a velocity of 25 m/s with a co-flow of air at 0.1 m/s and 300 K. The domain is a 2D square of 0.5 m^2 represented with 1500^2 grid points, and a jet diameter of 0.01 m. The simulation is performed within the Uintah Framework using the simulation software ARCHES. The explicit algorithm of ARCHES is limited to a maximum time-step size of $1e - 7$ s, reaching a physical time of 0.25 s. Figure 4.12a shows the original manifold contained in the ‘training’ data-set and Figure 4.12b shows the manifold calculated from the simulation data at $t = 0.25$ s, with the gray-scale representing s_{Z_2} . Here the x and y axis show Z_1 and Z_2 , respectively. It is observed that all of the points are bound to the manifold with no points leaving the observed space in the ‘training’ data. It is apparent that the Damköhler number of the system is large, as much of the data from the calculation is near the equilibrium solution.

The density (left) and temperature (right) fields at $t = 0.25$ s are shown in Figure 4.13, as well as Z_1 (left) and Z_2 (right), as well as their source terms in Figures 4.14 and 4.15. A small standoff distance of approximately the same size as the jet inlet (0.01 m) is observed. \mathbf{Z} is very insightful, as Z_1 is very similar to the Bilger's mixture fraction, in describing the degree of mixing between the premixed fuel stream and the surrounding air. Z_2 is very similar to the extent of reaction, showing the progress of the reaction as it proceeds to equilibrium conditions. Figures 4.16-4.20 show the species mass fractions fields computed from the model. It is evident from Figure 4.16 that most of the fuel has already been consumed before the jet exits the domain. Figure 4.17 shows good correlation between the

Table 4.5: *Nrms error* and R^2 statistics for the prediction of Φ for the laminar data-set while using pareto scaling and $q = 2$.

Φ	<i>nrms error</i>	R^2
H_2	0.05	0.998
O_2	0.02	0.999
O	0.07	0.996
OH	0.08	0.993
H_2O	0.07	0.996
H	0.18	0.973
HO_2	0.12	0.981
CO	0.01	1
CO_2	0.01	1
HCO	0.30	0.940
N_2	0.01	1
T	0.03	0.999
ρ	0.02	0.999
s_{Z_1}	0.11	0.988
s_{Z_2}	0.06	0.996

Table 4.6: Eigenvector matrix, A , from the PC analysis.

species weight	Z_1	Z_2	Z_3	Z_4	Z_5
H_2	1.771	-1.064	3.272	27.491	3.972
O_2	-1.831	-1.025	0.321	-0.297	2.331
H_2O	0.191	3.253	-12.323	1.087	3.414
CO	1.881	-0.679	0.131	-0.325	1.119
CO_2	0.224	-3.390	1.281	-0.246	1.459

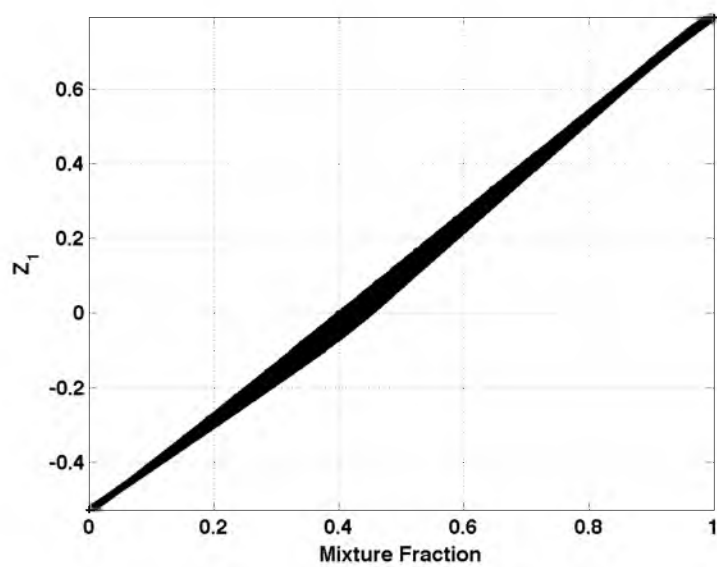


Figure 4.11: A scatter plot of mixture fraction (x-axis) versus Z_1 (y-axis), illustrating the correlation between the variables.

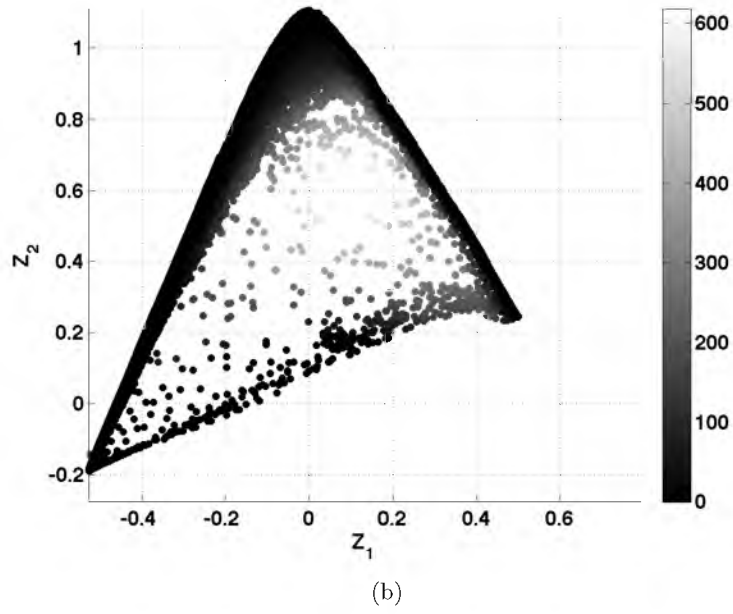
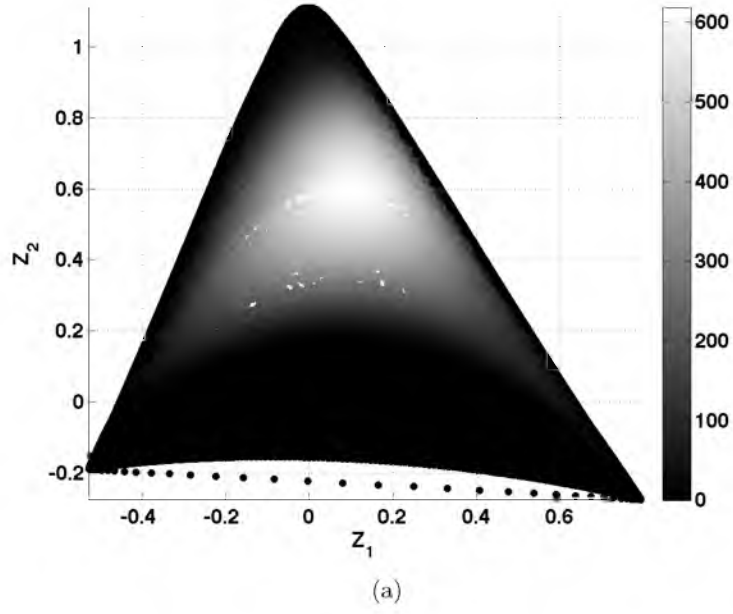


Figure 4.12: Scatter plots representing the $q = 2$ dimensional manifold with Z_1 on the x-axis, Z_2 on the y-axis, and the gray-scale representing s_{Z_1} . Plot (a) shows the representation of the manifold from the training data-set, and the plot (b) shows the represented manifold from the simulation results.

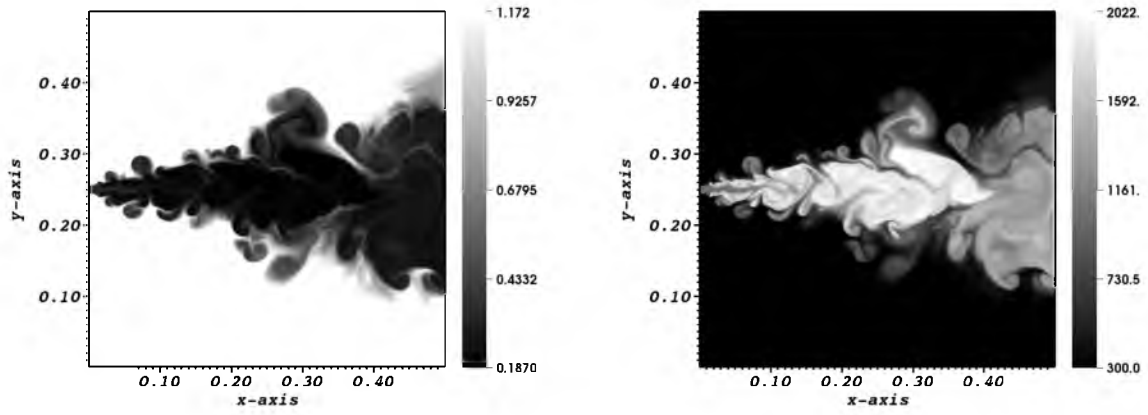


Figure 4.13: Density (left) and temperature (right) fields from the ARCHES calculation at $t = 0.25$ seconds.

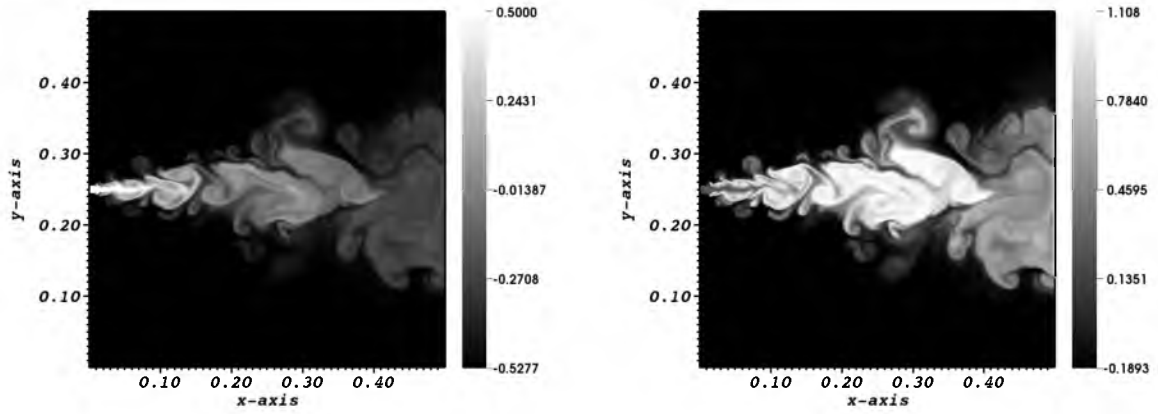


Figure 4.14: Z_1 (left) and Z_2 (right) fields from the ARCHES calculation at $t = 0.25$ seconds. Z_1 is highly correlated with Bilger's mixture fraction, and Z_2 is related to the extent of reaction being highly correlated with the temperature of the system.

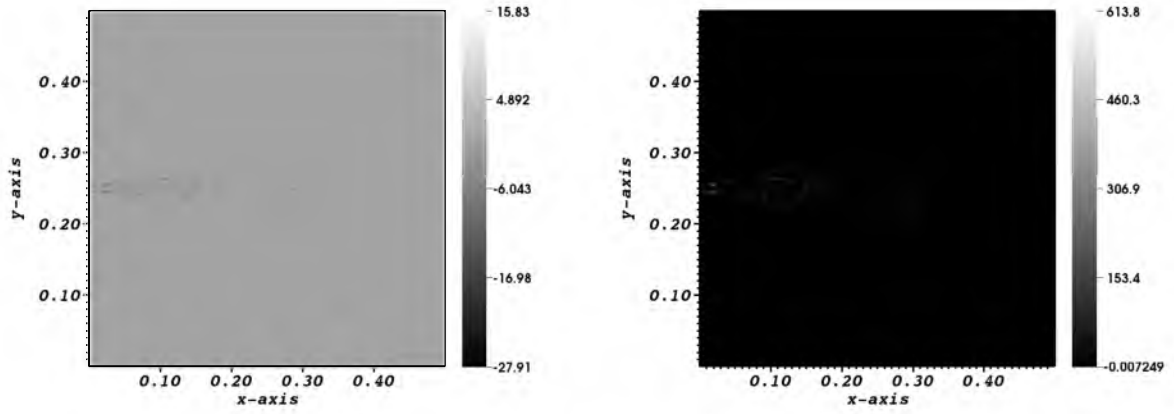


Figure 4.15: s_{Z_1} (left) and s_{Z_2} (right) fields from the ARCHES calculation at $t = 0.25$ seconds.

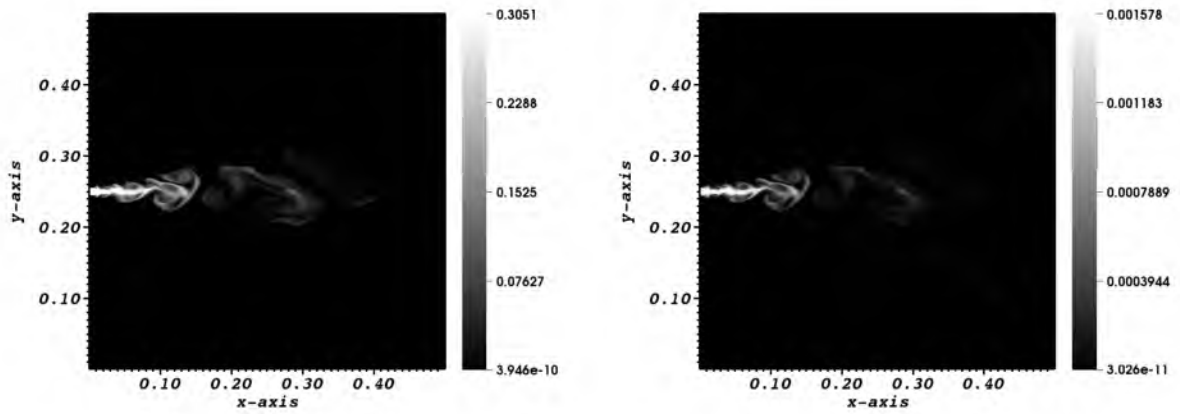


Figure 4.16: CO (left) and H_2 (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.

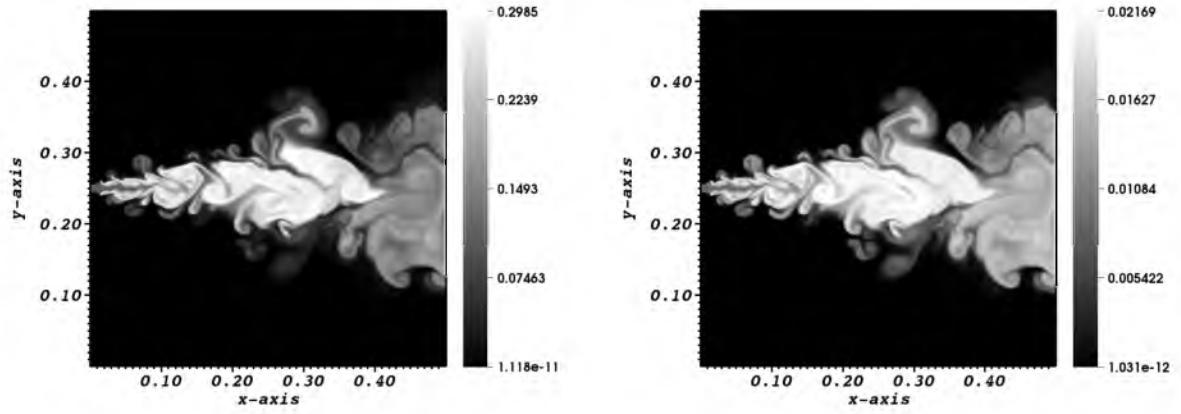


Figure 4.17: CO_2 (left) and H_2O (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.

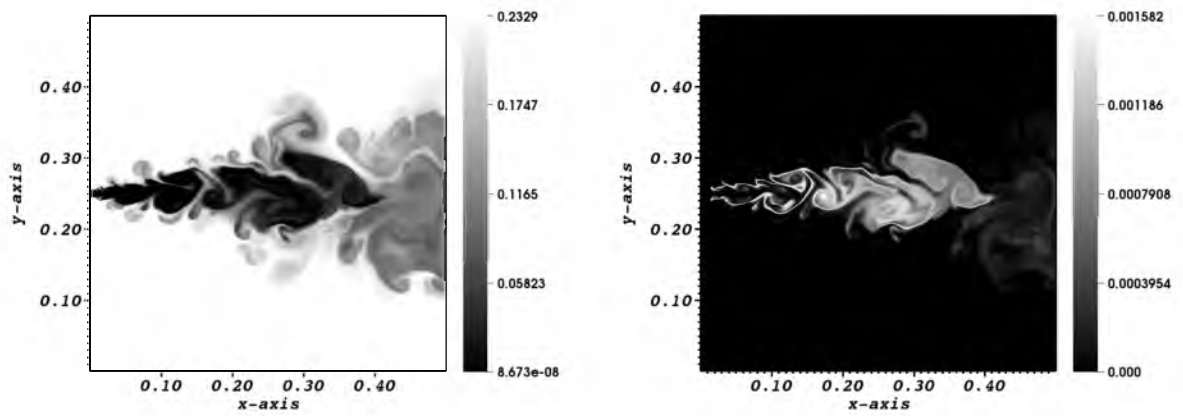


Figure 4.18: O_2 (left) and OH (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.

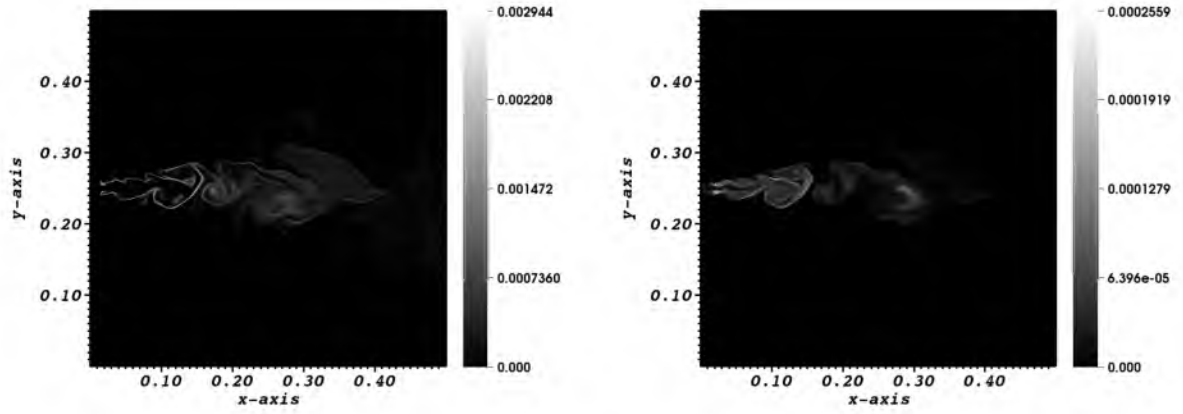


Figure 4.19: O (left) and H (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.

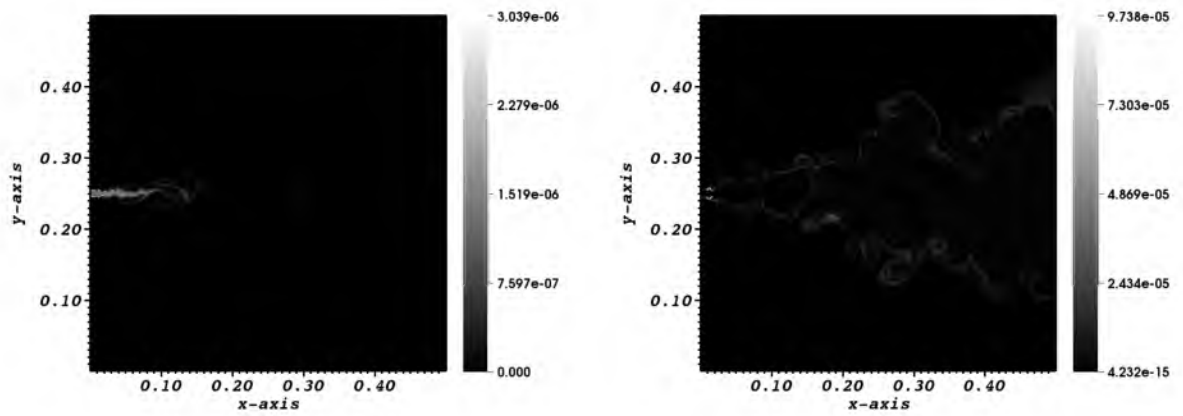


Figure 4.20: HCO (left) and HO_2 (right) species mass fraction fields from the ARCHES calculation at $t = 0.25$ seconds.

species products H_2O and CO_2 and the extent of reaction of the system.

4.6 Conclusion

The current work has addressed the ability to use nonlinear regression methods to estimate source-terms for the PC-transport combustion model. Various nonlinear regression methods have been analyzed showing the ability to produce accurate estimation even when using a lower number of \mathbf{Z} . In particular, the SVM and GPR methods have shown improved accuracy for estimating Φ . In addition, the effect of the various PCA-scaling methods on the regressibility of the system has been assessed. The pareto scaling method appears to achieve the greatest reduction with fewer components, and produces a highly regressible surface. The current work outlines an example of an *a priori* analysis which provides the best regression and scaling method for a given turbulent combustion data-set.

The work includes the first demonstrations of the PC-transport model using nonlinear regression within a numerical solver. In the case of the PSR, the model provided a computational reduction factor of 0.71, resulting in a very accurate representation of the original system with $q = 2$ variables of the 7 degrees of freedom in the system. The approach was demonstrated for the first time within a CFD solver. The CFD calculation can be considered a good proof of concept for the proposed approach. The results indicate that the simulation accesses part of the chemical state in the original training data without leaving the manifold.

4.7 Acknowledgments

We are grateful to our sponsor for which part of the present research was funded: The National Nuclear Security Administration under the Accelerating Development of Retrofittable CO_2 Capture Technologies through Predictivity program through DOE Cooperative Agreement DE NA 00 00 740.

CHAPTER 5

ADVANCING THE PCA APPROACH FOR COMBUSTION SYSTEMS

5.1 Introduction

In the current chapter, several new concepts are presented which seek to advance the PCA-combustion work given in the previous chapters, specifically for nonpremixed flows. Results from Chapter 5 have not been published; accordingly, the presentation will be less formal than what has been presented so far. The outline for this chapter is as follows:

- *Principal Component Constrained Equilibrium (PCCE)*

In Section 4.5.1.3, the PCA basis was calculated from a reduced-set of species mass fractions. The fact that the PCA basis is not calculated on the entire state-space can lead to issues, such as attempting to represent the other species mass fractions which were not used in the PC analysis, or not being able to adequately estimate other state variables such as temperature or density. Because of these issues, the PCCE framework was developed. Given the PC basis which represents the reduced-set of mass fractions, the remaining state-space variables are then estimated through equilibrium assumptions. If the mixture fraction is known the mass represented by the PCs and the total mass of the system are used to define the remaining mass which is allowed to reach equilibrium.

- *Mixture Fraction Conditioned Principal Component Analysis (MF-PCA)*

In Section 4.5.1.3, the identity of the first principal component Z_1 was highly correlated with the mixture fraction variable, ξ . Indeed, in the context of nonpremixed combustion, the mixture fraction variable is essential and intuitive, as the PC analysis confirms. Given that Z_1 from Section 4.5.1 has a nonzero reaction source-term s_{Z_1} , there seems to be an advantage in removing the variations in mixture fraction space from the system, doing the PC analysis on the conditioned data, and using mixture fraction to reconstruct the nonlinearity. Following this concept, the MF-PCA

framework was developed in-order to take full advantage of the systems nonlinearity described in mixture-fraction space.

5.2 PCCE

In order to increase the accuracy of the PC-Transport model, Chapter 4 discussed the ability to do the analysis on a reduced-set of variables. This concept leads to several issues which need to be addressed. The fact that the PC-basis is now optimal for only the major species does not guarantee a good representation of the minor species in the system, or the estimation of other state-space variables such as density and temperature. As discussed in the introduction, the basic idea behind PCCE is that while we represent a subset of chemical species using the PC-basis, the remaining variables can be estimated using equilibrium approximations. In order to describe the PCCE model, a simple theory section is now given, followed by a results section demonstrating the method on a simple strained flamelet solution of CH_4 .

5.2.1 PCCE theory

The general idea behind PCCE is that the majority of the mass in the system is being described by the PCs and evolves according to the source terms for the PCs. The small percentage of mass in the system, which in general represents minor and radical species, is assumed to evolve infinitely fast, and the conversion of this mass to species mass fractions can be fully described through equilibrium assumptions. Accordingly, the model name describes very well what is occurring: the principal components of the system are being used as a constraint, i.e. being held constant, during an equilibrium calculation which determines the final state of the minor species of the system. In other words, the equilibrium calculation of the small percentage of mass in the system is constrained by the large percentage of mass which is held constant during the equilibrium calculation.

Before discussing the constrained equilibrium calculation it is necessary to derive the equations which give the quantity of mass which is described by the PCs, and mass which is free to move toward equilibrium. In general, the i^{th} elemental mass fraction for a system containing E elements is given by the following equation:

$$K_i = \frac{\mathbf{Y} \mathbf{C}_i W_{e,i}}{\sum_{j=1}^E (\mathbf{Y} \mathbf{C}_j W_{e,j})}. \quad (5.1)$$

where \mathbf{Y} is row vector of species mass fractions, \mathbf{C}_i is a column vector describing the number of i atoms for each of the Q species. $W_{e,i}$ and $W_{e,j}$ are the atomic weights for element i , or

element j . For a system with a unity Lewis number, an estimation of the total elemental mass in the system is easily derived from the mixture fraction variable, ξ :

$$K_e^T = (K_e^{fuel} - K_e^{oxid}) \xi + K_e^{oxid} \quad (5.2)$$

Here the superscripts T , $fuel$, and $oxid$ represent the total, fuel stream, and oxidizer stream elemental mass fractions. It is clear from Equation 5.2 that K_e^T changes as a function of mixture fraction. Figure 5.1 shows K_e^T as a function of mixture fraction for a syngas system (0.5 H_2 , and 0.5 CO by volume) mixing with air.

As mentioned previously, the PC-transport equations in general will describe the evolution of the c major chemical species in the system. The i^{th} elemental mass fractions represented by PCs (\mathbf{Z}) is given by:

$$K_i^Z = \frac{\mathbf{Y}_c \mathbf{C}_{c,i} W_{e,i}}{\sum_{j=1}^E (\mathbf{Y}_c \mathbf{C}_{c,j} W_{e,j})}. \quad (5.3)$$

Here \mathbf{Y}_c is a row vector of the c chemical species mass fractions, used in the PC analysis, $\mathbf{C}_{c,i}$ is a column vector describing the number of i atoms for the c species in the system. Given a system without differential diffusion, it is now trivial to define the elemental mass in the system that is allowed to go to equilibrium:

$$K_e^{res} = K_e^T - K_e^Z \quad (5.4)$$

Here the superscript res , refers to the *residual* elemental mass that is not accounted for by the c species in the PCA analysis that is now assumed to proceed to equilibrium.

Now that the definition of the constrained elemental mass as well as the residual elemental mass have now been addressed, a discussion on how the equilibrium calculation is performed is given. The equilibrium described here refers to the minimization of Gibbs free energy which is carried out with a constant enthalpy, and constant pressure. In general, for a combustion system with low mach numbers, the pressure remains relatively constant; in addition, given the unity Lewis number assumption, the mass based enthalpy is a function of mixture fraction. Assuming the mixture is an ideal gas, the following equation describes the Gibbs function of the mixture [87]:

$$G_{mix} = \sum N_i \bar{g}_{i,T} = \sum N_i [\bar{g}_{i,T}^o + R_u T \ln (P_i/P)] \quad (5.5)$$

where N_i is the number of moles of species i , $\bar{g}_{i,T}$ and $\bar{g}_{i,T}^o$ are the Gibbs free energy of species i in the gas mixture at current conditions and at standard state, and P_i is the partial pressure of species i . In order to constrain the equilibrium calculation, a slight modification

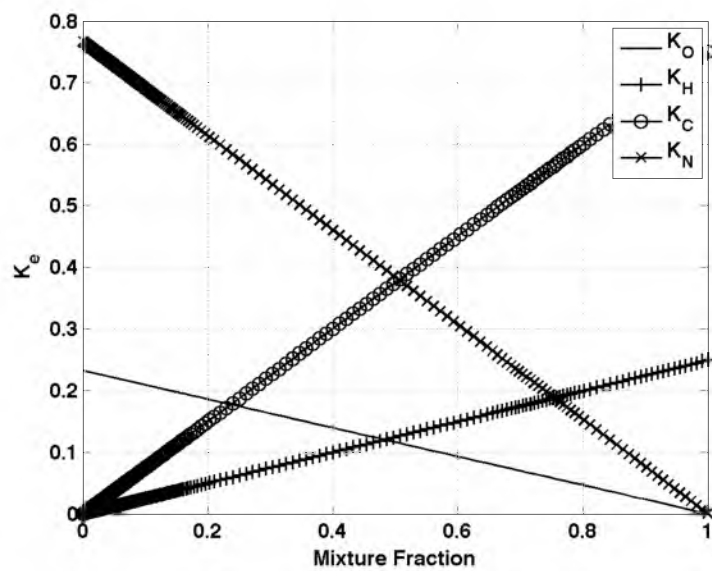


Figure 5.1: Total elemental mass fractions (K_e^T) for the flamelet data-set as a function of mixture fraction

of Equation 5.5 is required. Given a data-set with Q species and E elements, the system is now split into the c species which were used in the PC analysis and the $Q - c$ species which are determined by equilibrium assumptions. Now, a definition of the Gibbs function is as follows:

$$G_{mix} = \sum_{i \in c}^c N_i^{const.} [\bar{g}_{i,T}^\circ + R_u T \ln (P_i^{const.}/P)] + \sum_{i \notin c}^{Q-c} N_i [\bar{g}_{i,T}^\circ + R_u T \ln (P_i/P)] \quad (5.6)$$

Equation 5.6 shows that the Gibbs function is given by the contributions from the constrained species, which are held constant (indicated by the superscript *const.*), as well as the species which are free to go to equilibrium. The implementation of the above equation was done in Cantera [28]. Now, through the modified minimization of the Gibbs function, the equilibrium state can be estimated, giving an approximation for the $Q - c$ minor species in the system as well as the temperature and resultant density. The only inputs to the system are the species mass fractions for the c represented species, the enthalpy for a given mixture fraction, the pressure of the system, and K_e^{res} which are allowed to proceed to equilibrium.

5.2.2 PCCE results and conclusions

In order to demonstrate the approach, a simple 1D counter diffusion flame data-set was generated. The GRI3.0 mechanism was used to describe the reactions in the system [81], with the oxidizer being air (0.21 O_2 , and 0.79 N_2 by vol.) and the fuel being pure methane. The stoichiometric strain-rate of the flame, χ , was set to 0.1 1/s, with an initial temperature of 300 K. The computational package Cantera [28] was used to generate the data-set as well as perform the equilibrium calculations. The ordering for the constraints applied to the system was done according to the largest mean values for the data-set and are as follows for the first 12 chemical species: N_2 , CH_4 , H_2O , CO_2 , O_2 , CO , C_2H_2 , H_2 , OH , C_2H_4 , NO , and O . As an example, when $c = 4$, the $i \in c$ species are given as follows: $i = [N_2, CH_4, H_2O, CO_2]$.

Table 5.1 shows the fraction of elemental mass remaining while varying the number of constraints (c) on the system. With $c = 1$, the equilibrium solution of the system is obtained. As the first constraint was that of N_2 , nearly all of the elemental mass of N in the system was in the form of N_2 ; accordingly, Table 5.1 shows that virtually no N_2 remains for equilibrium. Upon addition of the 6th species, 5 percent or less of all of the elemental mass fractions remains for the equilibrium calculation. Figures 5.2-5.3 show K_e for the elements while varying c , confirming the results shown in Table 5.1. As c increases, K_e^Z

Table 5.1: Fraction of total elemental mass remaining $1 - \sum K_e^{res}$ while varying the number of constraints on the system.

c	O	H	C	N
1	1.0	1.0	1.0	0.0
2	1.0	0.22	0.22	0.0
3	0.58	0.05	0.22	0.0
4	0.31	0.05	0.11	0.0
5	0.11	0.05	0.11	0.0
6	0.0	0.05	0.02	0.0

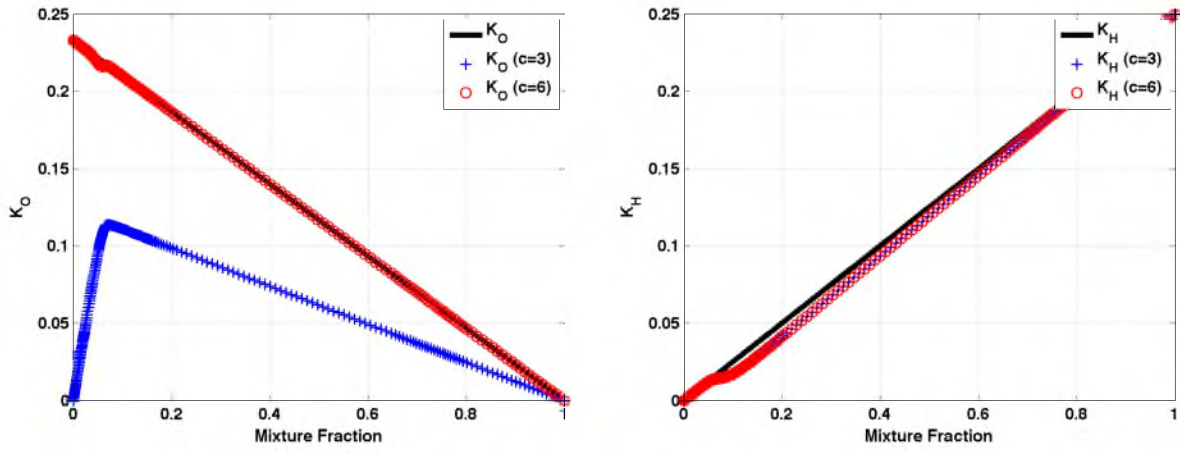


Figure 5.2: K_O (left) and K_H (right) elemental mass fractions as a function of mixture fraction while varying the number of constraints.

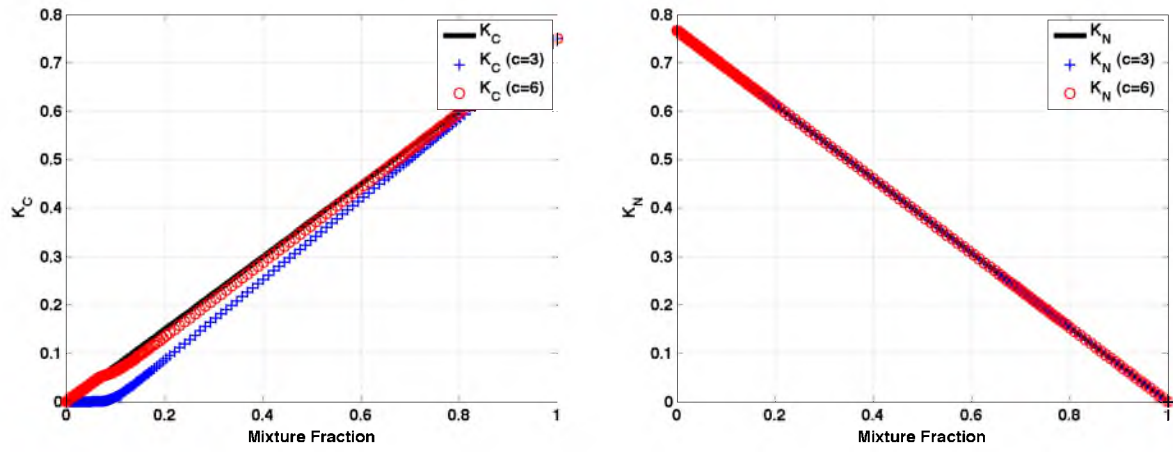


Figure 5.3: K_C (left) and K_N (right) elemental mass fractions as a function of mixture fraction while varying the number of constraints.

increases gradually while approaching K_e^T , leaving an ever-decreasing value for K_e^{res} . In other words, as more constraints are applied, the amount of mass described by Z increases, leaving less mass to be approximated from the equilibrium assumptions. Accordingly, one would assume an increase in accuracy as c increases.

The model is now applied to the data-set while varying c . For the results given here, it is assumed that the constrained species are represented perfectly from the regression of the PC basis, giving a better view of the error produced by using the equilibrium assumptions. Figure 5.4 shows the results for density (ρ with the units kg/m^3), and temperature (T in Kelvin) as a function of ξ while varying c . Only minor discrepancies are observed when $c > 6$ for both T and ρ . When examining the minor species in the system, which are not represented in the PC basis, a clear advantage is seen by using the model. Figure 5.5 shows the model approximation to CO (left) and H_2 (right) mass fractions while using 1 – 5 constraints for CO and 1 – 7 constraints for H_2 . It is evident that for CO , adequate representation is obtained with as few as 4 constraints, and 6 constraints are needed to reduce the error in H_2 . While looking at smaller species such as O and H , more discrepancies are seen, and as would be expected, a larger number of constraints are required to achieve higher accuracy. The draw back to adding constraints is the equilibrium calculation becomes more complicated because fewer species are available for equilibrium (as more constrained variables in the system are present). In addition, the more constraints used results in adding more nonlinearity to PC space, which in turn makes the nonlinear regression more difficult.

The preliminary results presented here indicate that PCCE has the potential for delivering highly accurate state-space variable such as temperature and density, with only a few constraints. Several of the minor species were also approximated with fewer than 7 constraints in the PCCE methodology. From this analysis, it is clear that PCCE can be very beneficial in its application to combustion systems.

5.3 MF-PCA

Turbulent combustion systems are inherently nonlinear. This is obvious when looking at the reaction-rates for chemical species, which are generally of the form:

$$R_i = -A \exp(-Ea/T) C_i^\alpha C_j^\beta \quad (5.7)$$

where A is the preexponential factor, Ea is the activation energy which has been scaled by the gas constant, T is the temperature, and C_i represents the molar concentration of species i . The large degree of coupling through net production rates from these reactions as well as the diffusion terms leave a highly complex nonlinear system which is extremely difficult

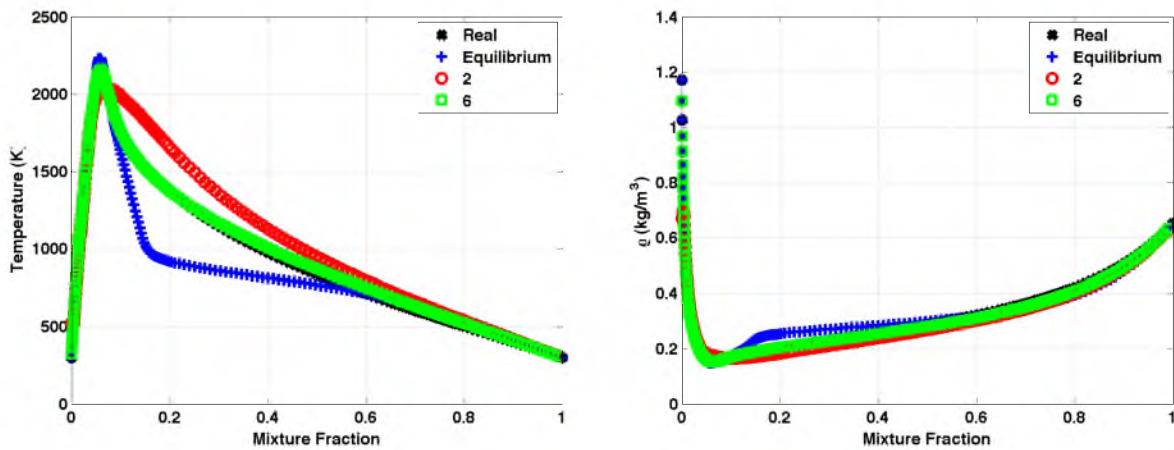


Figure 5.4: Temperature (left) and density (right) as a function of mixture fraction while varying the number of constraints.

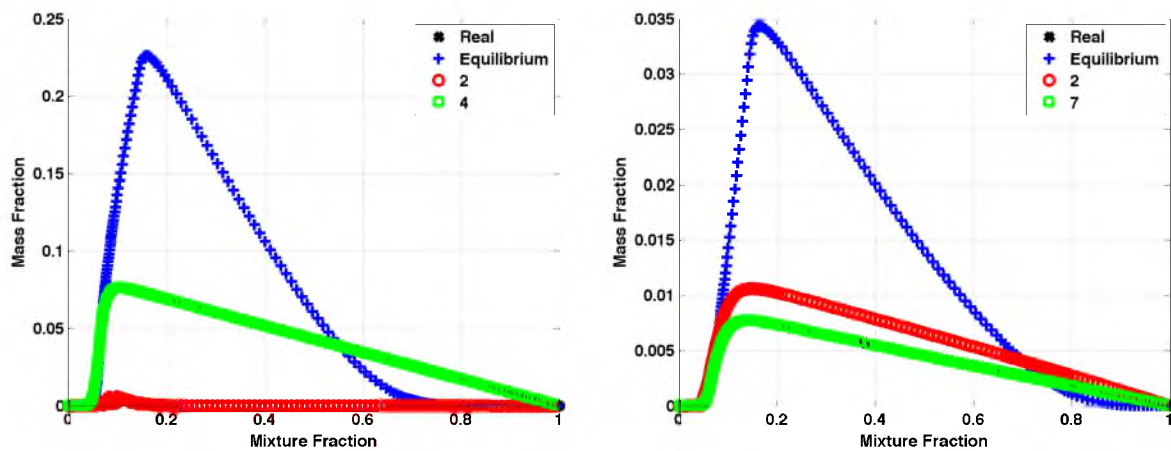


Figure 5.5: CO (left) and H_2 (right) mass fractions as a function of mixture fraction while varying the number of constraints.

to model. The PC-transport approach of Sutherland and Parente [85] seeks to represent this system with PCA, a linear model. Through nonlinear regression, the mapping of the state-space variables to the linear basis can be achieved; otherwise, a large number of PCs is required to adequately represent the state-space (see Section 4.5.1.2).

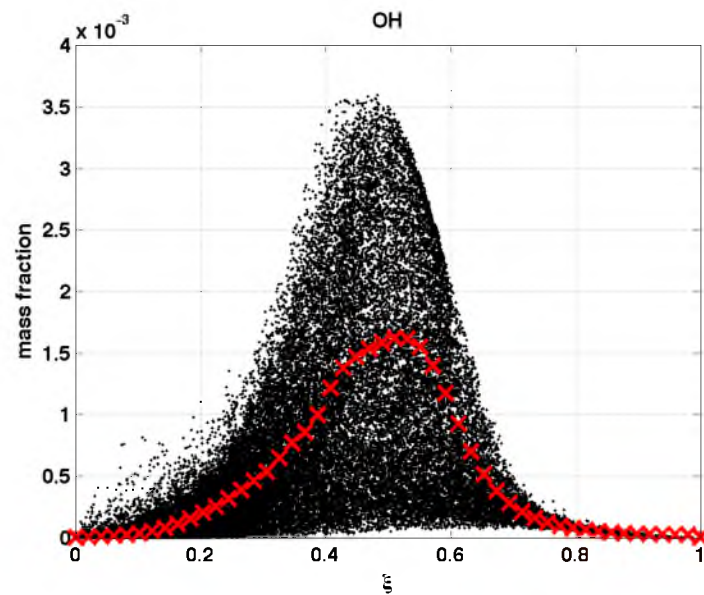
Chapter 4 showed an analysis of the basis matrix A of a nonpremixed syngas-jet data-set. In that system, Z_1 was highly correlated with mixture-fraction. This clearly indicates that the largest amounts of variation in the syngas data-set is described by the mixture fraction (see Figure 4.4). An attractive option for PCA would be to have a basis matrix \mathbf{A} which changes as a function of mixture fraction, thus providing a large degree of nonlinearity to the model. This concept was demonstrated in the work of Parente [62] for a nonpremixed DNS case where PCA was performed locally based on clustering the data in mixture fraction space. The local PC information dramatically increases the accuracy of the reconstruction. However, it is difficult to exploit this behavior within a combustion model, as the transport equations for the PCs are defined by the basis calculated from PCA, and changing the basis would change the identity of the transported variable.

The MF-PCA approach is a concept which has the potential of coupling the nonlinear effects seen in mixture fraction space with the PC analysis. The basic concept is to precondition the species mass fractions by subtracting the mixture-fraction dependent mean for each of the chemical species. This idea is illustrated in Figure 5.6a, which shows the radical species OH as a function of mixture fraction. The red markers in Figure 5.6a show the conditioning function which is dependent on mixture fraction and changes for each of the chemical species. Application of the conditioning leaves Figure 5.6b, which shows the conditioned OH radical with no variation in mixture fraction space. Now, the PC analysis is carried out on the conditioned data-set, with the understanding that the PCs are now attempting to represent a system which contains fewer nonlinearities.

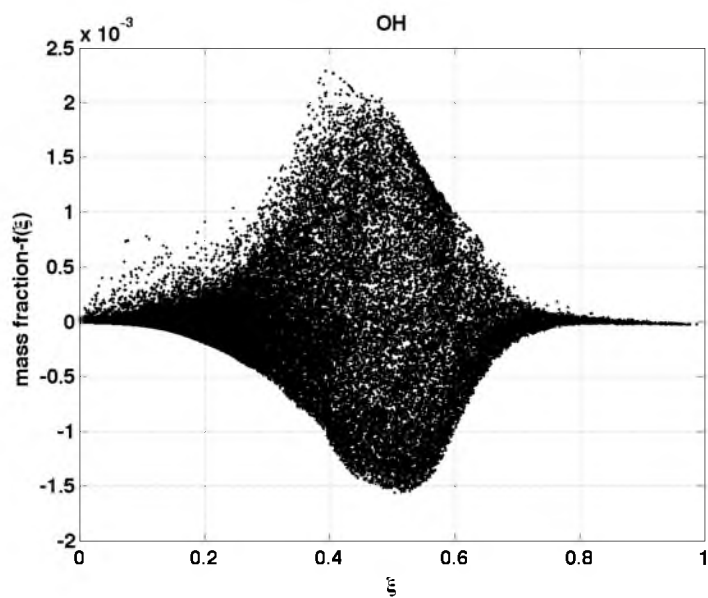
By conditioning the species mass fractions, the traditional transport equations defined by Sutherland and Parente [85] need to be reformulated. The following section describes the theory involved behind the conditioning, as well as the derivation for the new transport equations. In addition, a simple demonstration of the approach is presented on the same ODT data-set which was used in Chapter 4.

5.3.1 MF-PCA theory

Conditioning of the data requires an independent variable which can be used to map some function, f_i , to the i^{th} species mass fraction. Bilgers mixture fraction [6] is optimal for this and is defined as follows:



(a)



(b)

Figure 5.6: OH radical (a) and conditioned OH radical (b) as a function of mixture fraction, with red markers representing the condition function $f_{OH}(\xi)$.

$$\xi = \frac{2K_C/w_C + K_H/2w_H - 2(K_O - K_O^{oxid})w_o}{2K_C^{fuel}/w_C + K_H^{fuel}/2w_H + 2K_O^{oxid}w_o}. \quad (5.8)$$

Here the elemental mass fractions and molecular weights are the same as those which were described in Section 5.2. Mixture fraction is particularly interesting as it describes the change in chemical space from an oxidizer ($\xi = 0$) to a fuel ($\xi = 1$). Often much of the nonlinearity in the reacting species is observed when plotted in mixture fraction space.

Given a data-set with n observations, where each observation contains all of the species mass fractions, a value of ξ can be calculated for each observation. Figure 5.6a shows a plot of the n observations for the OH radical as a function of ξ . Using nonlinear regression, in this case, gaussian process regression (GPR) was used, a condition function is generated which returns the conditioned mean for species Y_i as a function of ξ :

$$Y_i^* = f_i(\xi) \quad (5.9)$$

Before applying PCA, the species mass fractions are conditioned by f_i , thus altering the definition of the PCs:

$$Z_q = \sum_{i=1}^{Ns} A_{i,q} (Y_i - f_i). \quad (5.10)$$

In order to generate transport equations for the new PCs, it is natural to start with the chemical species transport equations:

$$\frac{\partial \rho Y_i}{\partial t} + \rho u \cdot \nabla Y_i + \nabla \cdot (\rho D \nabla Y_i) = R_i \quad (5.11)$$

By conditioning Y_i and then projecting the PC basis, the following transport equation is found for Z_q :

$$\frac{\partial \rho Z_q}{\partial t} + \rho u \cdot \nabla Z_q + \nabla \cdot (\rho D \nabla Z_q) = \sum_{i=1}^{Ns} [A_{i,q} R_i] - \rho D \Delta \xi \sum_{i=1}^{Ns} \left[A_{i,q} \frac{\partial^2 f_i}{\partial \xi^2} \right] \quad (5.12)$$

The last term on the right-hand side of Equation 5.12 represents the source term of the projected conditioning function, or in other words, the source term for the underlying manifold. The term $\rho D \Delta \xi$ is the scalar dissipation rate. The term $\sum_{i=1}^{Ns} \left[A_{i,q} \frac{\partial^2 f_i}{\partial \xi^2} \right]$ has been referred to as the curvature of the manifold [72].

Equation 5.12 is a specific derivation for Z_q in the case that f_i is a function of ξ . If one desired to use alternative independent variables for creating f_i , a general solution to the transport equations for Z_q may be useful. In place of solving for an analytical expression

for the manifold source term, which resulted in Equation 5.12, the expression may be solved numerically:

$$\frac{\partial \rho Z_q}{\partial t} + \rho u \cdot \nabla Z_q + \nabla \cdot (\rho D \nabla Z_q) = \sum_{i=1}^{Ns} [A_{i,q} R_i] - \frac{D}{Dt} \left(\sum_{i=1}^{Ns} [A_{i,q} f_i] \right) \quad (5.13)$$

Projection of the conditioning function into PC space appears as follows:

$$Z_q^f = \sum_{i=1}^{Ns} [A_{i,q} f_i] \quad (5.14)$$

The last term in Equation 5.13 can literally be interpreted as the rate of change of the manifold in PC space, Z_q^f . In order to close Equation 5.13, an additional transport equation for Z_q^f is solved:

$$\frac{D}{Dt} (Z_q^f) = \frac{\partial \rho (Z_q^f)}{\partial t} + \rho u \cdot \nabla (Z_q^f) + \nabla \cdot (\rho D \nabla (Z_q^f)) \quad (5.15)$$

5.3.2 MF-PCA results and conclusions

Now, a simple *a priori* demonstration is given, testing the MF-PCA approach on the syngas ODT data-set referred to in Section 4.3 (please refer to Section 4.3 for a review of the pertinent details for this data-set).

First, the ability of the MF-PCA approach to reconstruct the species mass fractions is analyzed and compared with the standard PCA approach. Here the following equation is used to reconstruct the species mass fractions:

$$Y_i \approx \sum_{i=1}^{Ns} (A_{i,q} Z_q) + f_i(\xi) \quad (5.16)$$

Table 5.2 shows the *nrms* and R^2 values for the reconstruction of the species mass fractions using either the MF-PCA or the standard PCA approach. The approaches are compared assuming $q = 1$. It is only fair to state that the MF-PCA approach will also required mixture fraction, whereas the PC-transport approach has no such requirement. Because mixture fraction is a conserved scalar which changes very slowly compared to reacting chemical species, the comparison is fair, as the PCs always require a source-term. The results show that the reconstruction using the MF-PCA approach is significantly better for the chemical species. A similar pattern is to be expected for the reaction source-terms computed from the approximate species mass fractions. Table 5.2 also shows the error in reconstruction for s_{Z_1} , for either approach. Again, the MF-PCA model appears to have a much more accurate representation of the reaction rates.

Table 5.2: *nrms* and R^2 error for the reconstruction of the species mass fractions and s_{Z_1} using the standard PCA approach or MF-PCA

Y_i	R^2 (PCA)	R^2 (MF-PCA)	<i>nrms</i> (PCA)	<i>nrms</i> (MF-PCA)
H_2	< 0	0.941	1.026	0.244
O_2	0.997	1	0.054	0.015
O	< 0	0.943	1.032	0.239
OH	0.032	0.978	0.984	0.149
H_2O	0.885	0.970	0.339	0.174
H	0.531	0.953	0.685	0.216
HO_2	< 0	0.891	2.835	0.330
CO	0.883	0.999	0.342	0.025
CO_2	0.722	0.998	0.527	0.043
HCO	0.355	0.975	0.803	0.160
s_{Z_1}	0.086	0.807	0.956	0.440

In order to reduce the complexity of s_Z , PCA is now done on a reduced-set of species mass fractions (see Section 4.5.1.3). The scaled basis matrix for the new PCs is given in Table 5.3. With the nonlinear effects seen in mixture fraction space removed, Z_1 closely resembles the extent of reaction. The weights corresponding to the reactants: H_2 , O_2 , and CO , are negative with positive weights for the products: H_2O , and CO_2 .

Now, a comparison between the models where a reduced-set of species is used in the analysis can be made. Table 5.4 shows the reconstruction of the reduced-set of species for both of the approaches, where $q = 1$ for the MF-PCA, and standard PCA approaches. The MF-PCA again shows a significant improvement on the reconstruction.

Next, the ability to use nonlinear regression (GPR) for the entire set of state-space variables, $\Phi = \{Y_i, T, \rho, s_Z\}$, is checked and compared to what was reported in Section 4.5.1.3. Table 5.5 shows the approximation error for Φ using gaussian process regression with $q = 1$. In this case, the regression takes the form:

$$\Phi \approx f_{\Phi}(Z_q, \xi) \quad (5.17)$$

The results for the regression show good accuracy for all variables except for HO_2 . The results for s_{Z_1} are more accurate than what is found in the standard approach presented in Table 4.3. In addition, the results from Table 4.3 were a function of 2 principal components, again strengthening the case made for MF-PCA.

5.4 Conclusions

The results in this section demonstrate two promising concepts which seek to advance PCA-based combustion models.

PCCE has the potential of delivering highly accurate state-space variables in the case where PCA has been done on a reduced-set of chemical species. This approach may be useful in a scenario where nonlinear regression has difficulty in accurately mapping the species which were not represented in the analysis to the PC basis. The example given here demonstrates the models estimation of temperature, density, and several chemical species which were not constrained in the PC analysis. The model appeared to be very accurate on the data-set when up to 7 constraints were used.

MF-PCA has shown promise in its ability to add a nonlinear functionality to the original PCA method. In the demonstration given here, MF-PCA consistently decreased the error in approximation when compared to the standard PCA method. The major drawback of the approach is increased complexity in the transport equations. An analytical derivation was provided for the conditioned PC transport equations where the conditioning is a function

Table 5.3: Eigenvector matrix, A, from the PC analysis.

species weight	Z_1	Z_2	Z_3	Z_4	Z_5
H_2	-0.066	0.261	-0.022	-0.065	0.970
O_2	-0.469	0.247	0.080	-0.830	-0.154
H_2O	0.141	-0.881	0.119	-0.373	0.227
CO	-0.542	-0.251	-0.787	0.151	0.023
CO_2	0.680	0.179	-0.599	-0.381	-0.042

Table 5.4: $nrms$ and R^2 error for the reconstruction of a reduced-set of species mass fractions and s_{Z_1} using the standard PCA approach or MF-PCA. Here a comparison is made when $q = 1$ for both approaches.

Y_i	R^2 (PCA)	R^2 (MF-PCA)	$nrms$ (PCA)	$nrms$ (MF-PCA)
H_2	< 0	0.940	1.016	0.244
O_2	0.996	1	0.060	0.015
H_2O	0.882	0.970	0.343	0.174
CO	0.887	0.999	0.336	0.025
CO_2	0.715	0.998	0.534	0.043
s_{Z_1}	0.486	0.975	0.717	0.159

Table 5.5: $nrms$ error and R^2 statistics for the prediction of Φ while using pareto scaling and $q = 1$.

Φ	$nrms$ error	R^2
H_2	0.091	0.992
O_2	0.005	1
O	0.061	0.996
OH	0.080	0.994
H_2O	0.047	0.998
H	0.095	0.991
HO_2	0.246	0.938
CO	0.005	1
CO_2	0.009	1
HCO	0.064	0.996
N_2	0.005	1
T	0.021	1
ρ	0.026	0.999
s_{Z_1}	0.131	0.983

of mixture fraction. For generalization, a numerical derivation is given as well where the manifold source is computed by adding an additional transport equation.

CHAPTER 6

CONCLUSION

The preliminary research related to the use of principal component analysis in turbulent combustion systems has proven fruitful and merits continued study, development, and application within the field of turbulent combustion. The application of PCA to combustion has led to several important observations concerning inherent strengths and weaknesses of approach.

- *Strengths*

- PCA identifies orthogonal variables and gives a corresponding importance (eigenvalue) for each variable. In the case of turbulent combustion, it has been shown that the identified variables may have very rational definitions such as mixture fraction or extent of reaction. This is very attractive as one with no prior information about the system can rely on PCA to identify these fundamental variables and list them in order of importance.
- PCA is used to reduce the dimensionality of a system. This reduction applied to nonlinear data may be limited. However, in conjunction with nonlinear regression, a significant reduction can be achieved while maintaining reasonable accuracy. Depending on the computational resources and permissible error, the model can be tailored to the needs of the user.
- The PC-transport approach will most likely lead to a reduction in stiffness. This is evident because the leading PCs are generally highly correlated with larger species which are evolving more slowly in the system.

- *Weaknesses*

- A major draw-back to the approach is the fact that the PC basis is linear. Turbulent combustion systems are indeed nonlinear, thus the direct application of the approach may not result in a significant reduction. The development of

MF-PCA and MG-L-PCA attempt to reconcile this by adding some affect of nonlinearity to the system.

- Several disadvantages come from the fact that the approach requires data. For example, in order to use the PC-transport model with regression, a numerical data-set is required in order to resolve the source term issues.

Much of the work and development presented in the dissertation seeks to take advantage of the strengths from the approach and to avoid or overcome the weaknesses. An example of this can be seen in the MF-PCA approach which seeks to add nonlinearity to the system through preconditioning. Given a new perspective on the research that has been completed, several concepts are of interest for future work:

- The discussion on the Damköler number in Chapter 2 has opened up an interest in using the analysis to better inform which PCs are resolved (temporally, and physically) in the context of LES.
- The initial CFD results presented in Chapter 4 require validation and uncertainty quantification. The approach, until now, has only been demonstrated. Future work will include simulations which can easily be compared with experimental data.
- The MF-PCA approach has shown much promise. Future work may include implementation of the MF-PCA model in ARCHES.

Many of the ideas and concepts presented in this dissertation have come from the diligent efforts of students, committee members, and advisors. Using these resources as groundwork, several original contributions were presented in the dissertation:

- *Chemical Time-Scales*

Chapter 2 presented a novel technique for the calculation of chemical time-scales for turbulent combustion data. The approach identifies the limiting chemical time-scales in a system by utilizing a PCA-based analysis which identifies the principal variables in the system. With a clear identity for the key variables in the system, the source-term Jacobian approach is able to provide clear results free of dormant or redundant reaction time-scales. The approach has already been used in an analysis from another author [18].

- *Modified MG-PCA*

Chapter 3 presents a modification to the original MG-PCA and MG-L-PCA ap-

proaches. This modification utilizes the true PCs of the system instead of an approximation which was used previously. The effect of this modification increases the accuracy of the method, and in some cases, allows an additional reduction of 1 – 2 PVs.

- *Auto-ignition MG-PCA*

Chapter 3 also includes the first results for MG-PCA and MG-L-PCA in an unsteady constant pressure batch reactor. The results show a decrease in computational time, when compared to the full simulation. The MG-L-PCA approach reduced the required computation by 38%.

- *Regression Accuracy Study*

Chapter 4 provides more detail concerning the PC-transport approach. In particular, the ability to map the nonlinear state-space variables and source terms to the PC basis was assessed. It was shown that support vector regression, artificial neural network regression, and gaussian process regression provide superior results to previously demonstrated regression methods.

- *Perfectly Stirred Reactor Demonstration*

Chapter 4 shows the PC-transport approach using nonlinear regression within a numerical solver for the first time. The demonstration compares the full PSR results to a PSR simulation with 2 PCs using nonlinear regression, or 7 PCs using the standard approach. The results are very promising, showing the ability to predict various residence times with reasonable accuracy and reducing the computation by 75%.

- *Arches Implementation*

The PC-Transport approach is demonstrated for the first time within a CFD-solver. A 2D syngas case, where the model has been trained on a simple laminar dataset is demonstrated. The results are promising revealing several important physical phenomena such as flame stand-off and ignition.

- *PCCE*

Chapter 5 outlines the principal component constrained equilibrium approach. PCCE is demonstrated through *a priori* studies and shows the ability to estimate minor species when PCA is done on a reduced set of major species.

- *MF-PCA*

Chapter 5 introduces the mixture-fraction conditioned PCA model. The MF-PCA

model provides a distinct advantage in first conditioning the nonlinear data according to mixture fraction, and then using PCA.

The dissertation provides a step-forward in research as far as PCA-based combustion modelling is concerned. The dissertation also leaves many interesting future outlets for continued development and application.

REFERENCES

- [1] Numerical simulation of Delft-jet-in-hot-coflow (DJHC) flames using the eddy dissipation concept model for turbulence-chemistry interaction. *87*:537–567, 2011.
- [2] International Energy Agency. Fossil fuel energy consumption. *IEA Statistics OECD/IEA*, 2013.
- [3] J. Aminian, C. Galletti, S. Shahhosseini, and L. Tognotti. Key modeling issues in prediction of minor species in diluted-preheated combustion conditions. *Applied Thermal Engineering*, 31(16):3287 – 3300, 2011.
- [4] Javad Aminian, Chiara Galletti, Shahrokh Shahhosseini, and Leonardo Tognotti. Numerical investigation of a mild combustion burner: Analysis of mixing field, chemical kinetics and turbulence-chemistry interaction. *Flow, Turbulence and Combustion*, 88:597–623, 2012.
- [5] Amir Biglari and James C Sutherland. A filter-independent model identification technique for turbulent combustion modeling. *Combustion and Flame*, 2012.
- [6] RW Bilger. The structure of turbulent nonpremixed flames. In *Symposium (International) on Combustion*, volume 22, pages 475–488. Elsevier, 1989.
- [7] R.W. Bilger, S.H. Stårner, and R.J. Kee. On reduced mechanisms for methane-air combustion in nonpremixed flames. *Combustion and Flame*, 80:135–149, 1990.
- [8] A. Cavaliere and M. de Joannon. Mild combustion. *Progress in Energy and Combustion Science*, 30:329–366, 2004.
- [9] J. H. Chen, A. Choudhary, B. de Supinski, M. DeVries, E. R. Hawkes, S. Klasky, W. K. Liao, K. L. Ma, J. Mellor-Crummey, N. Podhorszki, R. Sankaran, S. Shende, and C. S. Yoo. Terascale direct numerical simulations of turbulent combustion using s3d. *Computational Science and Discovery*, 2, 2009.
- [10] F. C. Christo and B. B. Dally. Modelling turbulent reacting jets issuing into a hot and diluted coflow. *Combustion and Flame*, 142:117–129, 2005.
- [11] William S Cleveland, Eric Grosse, and William M Shyu. Local regression models. *Statistical models in S*, pages 309–376, 1992.
- [12] P. J. Coelho and N. Peters. Numerical simulation of a mild combustion burner. *Combustion and Flame*, 124:503–518, 2001.
- [13] I. Cohen, Q. Tian, X. Sean, Z. Thomas, and S. Huang. Feature selection using principal feature analysis. *Proc. of the 15th int. conf. on multimed.*, 15:301 – 304, 2007.

- [14] Scott D Cohen and Alan C Hindmarsh. CVODE, a stiff/nonstiff ODE solver in C. *Computers in physics*, 10(2):138–143, 1996.
- [15] A. Coussement, O. Gicquel, J. Caudal, B. Fiorina, and G Degrez. Three-dimensional boundary conditions for numerical simulations of reactive compressible flows with complex thermochemistry. *Journal of Computational Physics*, 231:5571 – 5611, 2012.
- [16] A. Coussement, O. Gicquel, and A. Parente. MG-local-PCA method for reduced order combustion modeling. *Proc. Combust. Inst.*, 34:1117 – 1123, 2013.
- [17] Axel Coussement, Olivier Gicquel, Benoit Fiorina, Gerard Degrez, and Nasser Darabiha. Multicomponent real gas 3-D-NSCBC for direct numerical simulation of reactive compressible viscous flows. *Journal of Computational Physics*, 245:259 – 280, 2013.
- [18] Axel Coussement, Benjamin Isaac, Olivier Gicquel, and Alessandro Parente. Assessment of different chemistry reduction methods based on principal component analysis: Comparison of the MG-PCA and score-PCA approaches. *Proc. Combust. Inst.*, 2014. Manuscript submitted for publication.
- [19] B. B. Dally, A. N. Karpetis, and R. S. Barlow. Structure of turbulent nonpremixed jet flames in a diluted hot coflow. *Proceedings of the Combustion Institute*, 29:1147–1154, 2002.
- [20] Scott G. Davis, Ameya V. Joshi, Hai Wang, and Fokion Egolfopoulos. An optimized kinetic model of h₂/co combustion. *Proc. Combust. Inst.*, 30:1283 – 1292, 2005.
- [21] Benoit Fiorina, Romain Baron, Olivier Gicquel, D Thevenin, Stéphane Carpentier, Nasser Darabiha, et al. Modelling non-adiabatic partially premixed flames using flame-prolongation of ILDM. *Combustion Theory and Modelling*, 7(3):449–470, 2003.
- [22] Benoit Fiorina, Olivier Gicquel, Stéphane Carpentier, and Nasser Darabiha. Validation of the FPI chemistry reduction method for diluted nonadiabatic premixed flames. *Combustion science and technology*, 176(5-6):785–797, 2004.
- [23] R. Fox. *Computational Models for Turbulent Reacting Flows*. Cambridge University Press, 2003.
- [24] Jerome H Friedman. Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67, 1991.
- [25] C. Galletti, A. Parente, and L. Tognotti. Numerical and experimental investigation of a mild combustion burner. *Combustion and Flame*, 151:649–664, 2007.
- [26] C. Gallletti, A. Parente, M. Derudi, R. Rota, and L. Tognotti. Numerical and experimental analysis of no emissions from a lab-scale burner fed with hydrogen-enriched fuels and operating in mild combustion. *International Journal of Hydrogen Energy*, 2009.
- [27] Olivier Gicquel, Nasser Darabiha, and Dominique Thévenin. Liminar premixed hydrogen/air counterflow flame simulations using flame prolongation of ILDM with differential diffusion. *Proceedings of the Combustion Institute*, 28(2):1901–1908, 2000.
- [28] DG Goodwin. An open-source, extensible software suite for CVD process simulation. *Chemical Vapor Deposition XVI and EUROCV*, 14(40):2003–08, 2003.

- [29] Evatt R Hawkes, Ramanan Sankaran, James C Sutherland, and Jacqueline H Chen. Scalar mixing in direct numerical simulations of temporally evolving plane jet flames with skeletal CO/H₂ kinetics. *Proceedings of the combustion institute*, 31(1):1633–1640, 2007.
- [30] Benjamin J. Isaac, Alessandro Parente, Chiara Galletti, Jeremy N. Thornock, Philip J. Smith, and Leonardo Tognotti. A novel methodology for chemical time scale evaluation with detailed chemical reaction kinetics. *Energy and Fuels*, 27:2255 – 2265, 2013.
- [31] J E Jackson. *A User’s Guide to Principal Component Analysis*. Wiley Series in Probability and Statistics, 1991.
- [32] I. T. Jolliffe. Discarding variables in a principal component analysis i: Artificial data. *Applied Statistics*, 21:160 – 173, 1972.
- [33] I. T. Jolliffe. *Principal Component Analysis*. Springer, New York, NY, 1986.
- [34] W.P. Jones and Rigopoulos Stelios. Rate-controlled constrained equilibrium: Formulation and application to nonpremixed laminar flames. *Combustion and Flame*, 142:223–234, 2005.
- [35] Nandakishore Kambhatla and Todd K Leen. Dimension reduction by local principal component analysis. *Neural Computation*, 9:1493–1516, 1997.
- [36] R.J. Kee, F.M. Rupley, E. Meeks, and J.A. Miller. Chemkin collection. *Release*, 3, 1996.
- [37] H. C. Keun, T. M. D. Ebbels, H. Antti, M. B. Bollard, O. Beckonert, B. Holmes, J. C. Lindon, and J. K. Nicholson. Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *Analytica Chimica Acta*, 490:265–276, 2003.
- [38] R. Acharya K.K. Kuo. *Fundamentals of Turbulent and Multiphase Combustion*. John Wiley and Sons, Inc., 2012.
- [39] W.J. Krzanowski. Selection of variables to preserve multivariate structure, using principal components. *Applied Statistics*, 36:22 – 33, 1987.
- [40] K.K. Kuo. *Principles of combustion*. A Wiley-Interscience publication. Wiley, 1986.
- [41] A. Liñán. The asymptotic structure of counterflow diffusion flames for large activation energies. *Acta Astronautica*, 1:1007–1039, 1974.
- [42] U. Maas and D Thévenin. Correlation analysis of direct numerical simulation data of turbulent non-premixed flames. *Proceedings of the Combustion Institute*, 27:1183–1189, 1998.
- [43] B. F. Magnussen. On the structure of turbulence and a generalized eddy dissipation concept for chemical reaction in turbulent flow. In *19th AIAA Aerospace Science Meeting*, January 1981.
- [44] B. F. Magnussen. Modeling of NO_x and soot formation by the eddy dissipation concept. In *Int. Flame Research Foundation, 1st topic Oriented Technical Meeting*, October 1989.

- [45] B. F. Magnussen. The eddy dissipation concept, a bridge between science and technology. In *Eccomas Thematic Conference on Computational Computation*, June 2005.
- [46] B.F. Magnussen and B.H. Hjertager. On mathematical modeling of turbulent combustion with special emphasis on soot formation and combustion. *Proceedings of the Combustion Institute*, 16:719–729, 1977.
- [47] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.
- [48] G. P. McCabe. Principal variables. *Technometrics*, 26:137 – 144, 1984.
- [49] Hessam Mirgolbabaei and Tarek Echekki. Nonlinear reduction of combustion composition space with kernel principal component analysis. *Combustion and Flame*, (In Press), 2013.
- [50] Hessam Mirgolbabaei and Tarek Echekki. A novel principal component analysis-based acceleration scheme for LES-ODT: An a priori study. *Combustion and Flame*, 160:898 – 908, 2013.
- [51] A. P. Morse. *Axisymmetric Turbulent Shear Flows with and without Swirl*. PhD thesis, London University, 1977.
- [52] Alireza Najafi-Yazdi, Benedicte Cuenot, and Luc Mongeau. Systematic definition of progress variables and intrinsically low-dimensional, flamelet generated manifolds for chemistry tabulation. *Combustion and Flame*, 159:1197 – 1204, 2012.
- [53] Duy Nguyen-Tuong, Matthias Seeger, and Jan Peters. Model learning with local gaussian process regression. *Advanced Robotics*, 23(15):2015–2034, 2009.
- [54] Isao Noda. Scaling techniques to enhance two-dimensional correlation spectra. *Journal of Molecular Structure*, 883-884:216 – 227, 2008.
- [55] Marcus Ó Conaire, Henry J. Curran, John M. Simmie, William J. Pitz, and Charles K. Westbrook. A comprehensive modeling study of hydrogen oxidation. *International Journal of Chemical Kinetics*, 36:603–622, 2004.
- [56] JA van Oijen and LPH de Goey. Modelling of premixed laminar flames using flamelet-generated manifolds. *Combustion Science and Technology*, 161(1):113–137, 2000.
- [57] Hsiao-Tien Pao. A comparison of neural network and multiple regression analysis in modeling capital structure. *Expert Systems with Applications*, 35(3):720–727, 2008.
- [58] A. Parente, C. Galletti, J. Riccardi, M. Schiavetti, and L. Tognotti. Experimental and numerical investigation of a micro-chp flameless unit. *Applied Energy*, 89(1):203 – 214, 2012.
- [59] A. Parente, C. Galletti, and L. Tognotti. Effect of the combustion model and kinetic mechanism on the mild combustion in an industrial burner fed with hydrogen enriched fuels. *International Journal of Hydrogen Energy*, 33:7553–7564, 2008.
- [60] A Parente, J C Sutherland, B B Dally, L Tognotti, and P J Smith. Investigation of the MILD combustion regime via principal component analysis. *Proceedings of the Combustion Institute*, 33(2):3333–3341, 2011.

- [61] A. Parente, J. C. Sutherland, P. J. Smith, and L. Tognotti. Identification of low-dimensional manifolds in turbulent flames. *Proc. Combust. Inst.*, 32:1579 – 1586, 2009.
- [62] Alessandro Parente. *Experimental and Numerical Investigation of Advanced Systems for Hydrogen-based fuel combustion*. PhD thesis, Università di Pisa, 2008.
- [63] Alessandro Parente and James C Sutherland. Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity. *Combustion and Flame*, 160:340 – 350, 2013.
- [64] T Passot and A Pouquet. Numerical simulation of compressible homogeneous flows in the turbulent regime. *J. Fluid Mech*, 181:441 – 466, 1987.
- [65] N. Peters. Laminar diffusion flamelet models in non-premixed turbulent combustion. *Progress in Energy and Combustion Science*, 10:319–339., 1984.
- [66] N. Peters. Laminar diffusion flamelet models in non-premixed turbulent combustion. *Progress in Energy and Combustion Science*, 10:319–339, 1984.
- [67] N. Peters. Laminar flamelet concepts in turbulent combustion. *Proceedings of the Combustion Institute*, 24:1231–1250, 1986.
- [68] N. Peters. The turbulent burning velocity for largescale and smallscale turbulence. *Journal of Fluid Mechanics*, 384:107–132, April 1999.
- [69] N. Peters. *Turbulent Combustion*. Cambridge University Press, 2004.
- [70] H. Pitsch and N. Peters. A consistent flamelet formulation for non-premixed combustion considering differential diffusion effects. *Combust. Flame*, 114:26–40, 1998.
- [71] T. Poinso and D. Veynante. *Theoretical and Numerical Combustion*. R.T. Edwards, Inc., 2001.
- [72] Stephen B Pope. Small scales, many species and the manifold challenges of turbulent combustion. *Proc. Combust. Inst.*, 34:1 – 31, 2013.
- [73] Naveen Punati, James C Sutherland, Alan R Kerstein, Evatt R Hawkes, and Jacqueline H Chen. An evaluation of the one-dimensional turbulence model: Comparison with direct numerical simulations of co/h₂ jets with extinction and reignition. *Proceedings of the Combustion Institute*, 33(1):1515–1522, 2011.
- [74] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011.
- [75] Carl Edward Rasmussen. Gaussian processes for machine learning. 2006.
- [76] R Rawat, J Spinti, S Kumar, S Borodai, and A Violi. Large eddy simulations of accidental fires using massively parallel computers. 2003.
- [77] Rajesh Rawat, Jennifer P Spinti, Wing Yee, and Philip J Smith. Parallelization of a large scale hydrocarbon pool fire in the uintah pse. In *ASME 2002 International Mechanical Engineering Congress and Exposition*, pages 49–55. American Society of Mechanical Engineers, 2002.
- [78] M. Rehm, P. Seifert, and B. Meyer. *Computers and Chemical Engineering*, 33(2):402 – 407, 2009.

- [79] J.H. Frank R.S. Barlow. *Proc. Combust. Inst.*, 27:1087–1095, 1998.
- [80] R.S. R.S. Barlow, G.J. Fiechtner, C.D. Carter, and J.Y. Chen. *Combust. Flame*, 120:549–569, 2000.
- [81] GP Smith, DM Golden, M Frenklach, NW Moriarty, B Eiteneer, M Goldenberg, CT Bowman, R Hanson, S Song, WC Gardiner Jr, et al. Gri-mechanism 3.0. 1999.
- [82] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.
- [83] S.Rigopoulos. The rate-controlled constrained equilibrium RCCE method for reducing chemical kinetics in systems with time-scale separation. *International Journal for Multiscale Computational Engineering*, 5(1):11–18, 2007.
- [84] H. Steinhaus. Sur la division des corp materiels en parties. *Bulletin L’Academie Polonaise des Science*, 4:801 – 804, 1956.
- [85] J. Sutherland and A. Parente. Combustion modeling using principal component analysis. *Proc. Combust. Inst.*, 32:1563–1570, 2009.
- [86] J.C. Sutherland. *Evaluation of mixing and reaction models for large-eddy simulation of nonpremixed combustion using direct numerical simulation*. PhD thesis, Department of Chemical and Fuels Engineering, University of Utah, 2004.
- [87] Stephen R Turns et al. *An introduction to combustion*, volume 499. McGraw-Hill New York, 1996.
- [88] J. A. van Oijen. Flamelet-generated manifolds: Development and application to premixed flames. *Eindhoven University of Technology*, 2002.
- [89] JA Van Oijen and LPH De Goey. Modelling of premixed counterflow flames using the flamelet-generated manifold method. *Combustion Theory and Modelling*, 6(3):463–478, 2002.
- [90] D. Veynante and L. Vervisch. Turbulent combustion modeling. *Progress in Energy and Combustion Science*, 28:193–266, 2002.
- [91] Yue Yang, Stephen B Pope, and Jacqueline H Chen. Empirical low-dimensional manifolds in composition space. *Combustion and Flame*, 160:1967 – 1980, 2013.
- [92] RA Yetter, FL Dryer, and H. Rabitz. A comprehensive reaction mechanism for carbon monoxide/hydrogen/oxygen kinetics. *Combustion Science and Technology*, 79(1-3):97–128, 1991.